PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
A61K 49/00, C12P 21/06, C12N 5/00, 15/00, C07H 19/00, 21/04 — A1

(11) International Publication Number: WO 98/03206

(43) International Publication Date: 29 January 1998 (29.01.98)

(21) International Application Number: PCT/US97/12933

(22) International Filing Date: 24 July 1997 (24.07.97)

(30) Priority Data:

60/022,563	24 July 1996 (24.07.96)	US
60/025,006	20 August 1996 (20.08.96)	US

(71) Applicant *(for all designated States except US)*: UNIVERSITY OF MARYLAND [US/US]; Office of Technology Liaison, 34122 Knox Road, College Park, MD 20742 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: MATHER, Ian, H. [US/US]; 9119 St. Andrews Place, College Park, MD 20740 (US). OGG, Sherry, L. [US/US]; 905 Briggs Chaney Road, Silver Spring, MD 20905 (US). JACK, Lucinda, J., W. [GB/US]; 11912 Hall Shop Road, Clarksville, MD 21029 (US).

(74) Agent: POULOS, James, A., III; Watson Cole Stevens Davis, P.L.L.C., Suite 1000, 1400 K Street N.W., Washington, DC 20005-2477 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: THE BUTYROPHILIN GENE PROMOTER AND USES THEREOF

(57) Abstract

The DNA sequence of the mouse butyrophilin gene and its promoter is disclosed and analyzed. In addition, expression of the mouse butyrophilin gene is characterized. Further, use of the butyrophilin promoter for expressing polypeptides in the milk of a transgenic animal and for screening substances for carcinogenicity is disclosed.

Mouse Viral Promoter

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

THE BUTYROPHILIN GENE PROMOTER AND USES THEREOF

FIELD OF THE INVENTION

The present invention relates generally to the butyrophilin gene promoter. More specifically the present invention relates to the use of the butyrophilin gene promoter for the production of heterologous proteins in the milk of transgenic animals and for the detection of carcinogenic substances. Applicants hereby incorporate by reference the subject matter of U.S. Serial No. 60/022,563.

BACKGROUND OF THE INVENTION

Butyrophilin is the major integral protein associated with the fat-globule membrane (FGM) in the milk of many species and is believed to play a role in the mechanism of milk secretion. See Franke *et al.*, *J. Cell Biol.* 89: 485-494 (1981); Jack and Mather, *J. Biol. Chem.* 265: 14481-14486 (1990); and Jack and Mather, *J. Dairy Sci.* 76: 3832-3850 (1993); each of which is herein incorporated by reference. Expressed on the apical surfaces of mammary epithelial cells, butyrophilin is a type I glycoprotein, comprising a glycosylated exoplasmic domain, a membrane anchor approximately in the middle of the sequence, and a long cytoplasmic tail.

Butyrophilin is a member of the immunoglobulin superfamily (IgSF) (Gardinier *et al.*, *J. Neurosci. Res.* 33, 177-187 (1992)), with closest *structural* homology in the exoplasmic domain to the B7.1 (CD 80) and B7.2 (CD 86) receptors (Linsley *et al.*, *Protein Sci.* 3: 1341-1343 (1994)). Hallmarks of these proteins are two exoplasmic immunoglobulin-like domains;

one of the variable (V) or intermediate (I) type (Williams and Barclay, *Ann. Rev. Immunol.* 6, 381-405 (1988); Harpaz and Chothia, *J. Mol. Biol.* 238, 528-539 (1994)) close to the N-terminus; and one of the constant (C) type (Williams and Barclay, 1988) close to the membrane anchor. Other proteins that are homologous with butyrophilin in the exoplasmic domain include myelin oligodendrocyte glycoprotein (MOG), a component of the myelin sheath (Gardinier *et al.*, 1992), and the chicken B-G antigens associated with the avian major histocompatibility complex (Miller *et al., Proc. Natl. Acad. Sci. U.S.A.* 88: 4377-4381 (1991)). MOG and the B-G antigens have shorter exoplasmic domains with one V-set immunoglobulin-like fold (Gardinier *et al.*, 1992; Miller *et al.*, 1991). The inclusion of butyrophilin in the IgSF and the B-G antigen system suggests that butyrophilin has immune functions.

The C-terminal cytoplasmic domain of butyrophilin is similar to the C-termini of a group of proteins that contain zinc finger and coiled-coil domains. These proteins may bind nucleic acids or proteins (Bellini *et al., J. Cell Biol.* 131: 563-570 (1995)) and include *ret* finger protein (RFP) (Takahashi *et al., Mol. Cell Biol.* 8: 1853-1856 (1988)), nuclear antigen A of Sjögren's syndrome (SSA/Ro) (Chan *et al., J. Clin. Invest.* 87: 68-76 (1991)), Xenopus nuclear factor 7 (XNF7) (Reddy *et al., Develop. Biol.* 148: 107-116 (1991)), PwA33 from *Pleurodeles waltl* (Bellini *et al., EMBO J.* 12: 107-114 (1993)), and acid finger protein (AFP) (Chu *et al., Genomics* 29:229-239 (1995)). At the DNA level, this homologous region encompasses an exon, named B30.2, which was mapped together with the MOG, RFP and butyrophilin genes to the human MHC class I region of chromosome 6 (Vernet *et al., J. Mol. Evol.* 37: 600-612 (1993)). Based on these observations, Vernet *et al.* (1993) suggested that the butyrophilin gene evolved in the MHC by the shuffling of exons between an ancestral MOG

gene which gave rise to the exon encoding the I-set immunoglobulin-like domain o butyrophilin, and an ancestral RFP gene, which gave rise to the B30.2 region of the butyrophilin gene.

Butyrophilin is specifically expressed in mammary tissue, with expression being maximal during lactation. This mammary-specific expression of the butyrophilin gene is assumed to be under the control of the butyrophilin promoter. Since butyrophilin constitutes a significant portion of the total protein associated with the milk FGM of many species, *i.e.*, more than 40% of the total FGM-associated protein in bovine milk is butyrophilin, the butyrophilin promoter is an attractive mammary-specific promoter for producing heterologous protein in the milk of transgenic mammals.

Promoters of other mammary-specific genes, *i.e.* the casein, whey acidic protein, α-lactalbumin, and β-lactoglobulin genes, have been used to direct the production of foreign proteins in the milk of transgenic animals. Recent analysis of these mammary-specific gene promoters has led to the identification of a number of potentially important regulatory elements which mediate the lactogenic response. These elements include binding sites for the following: CTF/NF1 in the β-lactoglobulin (Watson *et al.*, *Nucl. Acids Res.* 19: 6603-6610 (1991)) and whey acidic protein genes (Li and Rosen, *Mol. Cell Biol.* 15: 2063-2070 (1995)); Oct 1 in the bovine α_{s2}-casein gene (Groenen *et al.*, *Nuc. Acids Res.* 20: 4311-4318 (1992)); a single-stranded nucleic acid binding protein which negatively regulates the β-casein gene (Altiok and Groner, *Mol. Cell Biol.* 14: 6004-6012 (1994)); Ets-related proteins which stimulate (Welte *et al.*, *Eur. J. Biochem.* 223: 997-1006 (1994)), and unidentified factor(s) which negatively regulate, the whey acidic protein gene (Kolb *et al.*, *J. Cellul. Biochem.* 56: 245-261 (1994)), and a pregnancy-specific protein which modulates progesterone-mediated repression of the

mouse β-casein gene (Lee and Oka, *J. Biol. Chem.* 267: 5797-5801 (1992)). Several genes, including the most intensively studied rodent β-casein gene promoters, contain C/EBP (Doppler *et al.*, *J. Biol. Chem.* 270: 17962-17969 (1995); Raught *et al.*, *Molec. Endocrinol.* 9: 1223-1232 (1995)), YYI (Meier and Groner, *Mol. Cell Biol.* 14: 128-137 (1994); Raught *et al.*, *Mol. Cell Biol.* 14: 1752-1763 (1994)), MGF/STAT5 (Watson *et al.*, *Nucl. Acids Res.* 19: 6603-6610 (1991); Groenen *et al.*, 1992; Wakao *et al.*, *EMBO J.* 13: 2182-2191 (1994)) and glucocorticoid response elements (Raught *et al.*, 1995). In addition, the promoter of the housekeeping gene, β1,4-galactosyltransferase, contains binding sites for AP-2 and CTF/NF1, which regulate the synthesis of a mammary-specific 3.9 kb transcript (Rajput *et al.*, *J. Biol. Chem.* 271:5131-5142 (1996)).

The basis for mammary-specific expression is poorly understood in any system. A so-called "milk-box" sequence, first identified in the proximal α-lactalbumin gene promoter in several species, is also conserved in many of the casein genes (Laird *et al.*, *Biochem. J.* 254: 85-94 (1988)), and encompasses binding sites for YY1, STAT5 (Meier and Groner, 1994; Raught *et al.*, 1994) and C/EBP isoforms (Doppler *et al.*, 1995; Raught *et al.*,1995). Also there are three conserved sequences in the casein genes referred to as blocks A, B, and C (Yoshimura, M. and Oka, T., Gene 78, 267-275). Raught *et al.* (1995) have recently suggested that casein gene expression is regulated by composite response elements (CoREs) comprising STAT5 and glucocorticoid response elements and C/EBP binding sites.

For the first time, the inventors have cloned and sequenced the mouse butyrophilin gene, including its promoter region and have found that the promoter sequence has no significant similarities with the published sequences of these other mammary-specific promoters.

Analysis of the butyrophilin promoter sequence showed that the butyrophilin promoter contains many potential regulatory elements associated with immune system genes including α- and γ-interferon response elements, and consensus sequences for TCF-1 and PU.1. (PU.1 is a macrophage and B cell-specific transcription factor related to the *ets* oncogene. *See* Klemsz, *et al.*, *Cell* 61: 113-125 (1990)). In addition, the inventors have found that the proximal region of the butyrophilin promoter contains a repeat element of three granulocyte-macrophage colony-stimulating factor (GMCSF) sites which in the same context has been shown to regulate the mitogen-inducible expression of GMCSF in T cells (Nimer *et al.*, *Mol. Cell. Biol.* 10: 6084-6088 (1990), herein incorporated by reference). Thus, the butyrophilin promoter is also useful for the detection of carcinogenic substances.

BRIEF DESCRIPTION OF THE INVENTION

The present invention provides the sequence of the 5′ flanking region and transcriptional unit of the mouse butyrophilin gene (*Btn*). In particular, it provides the Btn promoter and transcriptional regulatory elements contained therein.

Accordingly, an object of the invention is an isolated and purified DNA fragment comprising a DNA sequence encoding a polypeptide having the biological activity of a butyrophilin protein.

Another object of the invention is an isolated and purified DNA fragment comprising a DNA sequence having the biological activity of a butyrophilin promoter.

An additional object of the present invention is a rDNA construct for expressing a polypeptide in the mammary gland of a mammal. The rDNA construct comprises a butyrophilin promoter operatively linked to the DNA sequence encoding a desired polypeptide.

The rDNA construct may also have a DNA sequence encoding a signal sequence operatively linked to the DNA sequence encoding the polypeptide. Preferably, the signal sequence is a milk protein signal sequence. The DNA construct may also include the transcriptional unit and/or 3′ flanking sequence of the butyrophilin gene.

It is a further object of this invention to provide a transgenic animal which produces a desired polypeptide in its mammary gland. This is achieved by introducing a rDNA construct comprising a butyrophilin promoter operatively linked to the DNA sequence encoding the polypeptide into at least the mammary epithelial cells of the mammal. Alternatively, the rDNA construct may be introduced into a germ line of a mammal, thus subsequent generations will also express the desired polypeptide in their milk.

Another aspect of the present invention is the use of the mitogen-inducible elements in the butyrophilin promoter to detect mitogenic properties of potential carcinogens from a variety of sources. For example, substances found in the environment or isolated from food sources could be tested for carcinogenicity. The mitogenic properties of a substance are assessed by detecting activation of the butyrophilin promoter in cells exposed to the substance, either by detection of butyrophilin mRNA or protein, or by detecting expression of a reporter gene under the control of a butyrophilin promoter.

Yet another object of the invention is diagnosis of disease states such as breast cancer by screening mammary and nonmammary tissues of nonlactating animals for the expression of butyrophilin.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1A is a schematic representation of the λBtn1 clone, showing the location of subclones prepared from λBtn1 which were used to generate the sequence of the mouse butyrophilin gene and 5′ flanking region.

FIG. 1B is a schematic representation of the structure of the mouse Btn gene, showing the location of the exons and introns.

FIG. 1C is a schematic representation of mouse butyrophilin cDNA, showing the location of the cDNA subclones used to sequence mouse butyrophilin cDNA.

FIG. 2 A-C shows the location of putative regulatory elements in the proximal 5′ flanking region of the mouse butyrophilin gene.

DETAILED DESCRIPTION OF THE INVENTION

The invention provides the sequence of the mouse butyrophilin gene, and approximately 4.6 kb sequence of its 5′ flanking region, which is also referred to as the butyrophilin promoter. These sequences were obtained from a clone isolated from a murine genomic library as described below.

Example 1: Cloning the Mouse Butyrophilin Gene

Screening of genomic library and cloning of λBtn1. A 129 ES cell genomic library in Lambda DASH® (Stratagene, La Jolla, CA) was screened with a 2.3 kb *Xho*I-*Xba*I fragment of cDNA encoding bovine butyrophilin (Jack and Mather, 1990). Plaque DNA (total of 500,000 pfu) was transferred to nylon membranes (Dupont, Boston, MA), denatured in 0.5N NaOH, neutralized and cross-linked to the membranes by exposure to ultraviolet light using a UV Stratalinker® 1800 (Stratagene). Membranes were incubated for 2 h at 42°C in prewash

solution (Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press, NY, (1989)), followed by 6 h at 42° C in pre-hybridization solution, and were then incubated overnight at 42°C in hybridization solution with the bovine butyrophilin cDNA fragment which had been labelled with [α-^{32}P]-dCTP to a specific activity of 10^9 cpm/μg by the random priming method (Feinberg and Vogelstein, *Anal. Biochem.* 132: 6-13 (1983)). Filters were briefly rinsed with 2X SSC (Sambrook *et al.*, 1989) and then washed three times at 55° C with 2X SSC containing 0.1 % (w/v) SDS, for 20 min each time, and the cDNA bound to the membranes detected by exposure to X-ray film, overnight, at -80° C. One potentially positive plaque was detected from a total of 500,000 pfu's screened and this cloned DNA was designated λBtn1, which has been deposited with the American Type Culture Collection as ATCC designation 97513.

To confirm that λBtn1 contained the mouse butyrophilin gene, samples of the cloned DNA (λBtn1), mouse, and bovine genomic DNA were digested with either *Eco*R1 or *Hind*III and subjected to Southern analysis using the [^{32}P]-labelled 2.3kb *Xba*I fragment of heterologous bovine cDNA as a probe.. In each case, digestion with the restriction endonucleases generated a characteristic pattern of DNA fragments which hybridized to the [^{32}P]-labelled probe. Similar patterns of radiolabelled bands were detected in the genomic DNA and λBtn1 samples (data not shown). As the sequence of λBtn1 became available, a mouse cDNA probe, mcDNA3, encoding the 3′ end of exon 3 through the first 396 bp of exon 7 (see FIG. 1C) was prepared by RT-PCR. A Southern blot, similar to those described above and probed with this homologous cDNA, confirmed the identity of λBtn1 (data not shown).

Example 2: Sequencing the Mouse Butyrophilin Gene

Subclones spanning over 14 kb of λBtn1 DNA were prepared (see FIG. 1A) and sequenced on both strands using the *fmol* sequencing kit from Promega Corp (Madison, WI). Autoradiographs were scanned with a Molecular Dynamics Computing Densitometer and the sequences read using the Image Quant®, Version 3.30 software package (Molecular Dynamics, Sunnyvale, CA). The computer program MACAW (Schuler *et al.*, 1991) was used to compile the full-length sequence from the sequencing gels and the sequence is shown in SEQ ID NO:1. The entire *Btn* sequence has been deposited in the GenBank Data Base under Accession No. U67065. The butyrophilin promoter is contained within the first 4,693 nucleotides of SEQ ID NO:1. The proximal part (first 1750 nucleotides) of this region is shown in SEQ ID NO: 2 and schematically illustrated in FIG. 2, with the nucleotides being renumbered in conventional format, *i.e.*, where the most proximal transcriptional start site (see below) is designated +1.

Example 3: Expression of the Mouse Butyrophilin Gene

Mapping the 5′ end of mouse butyrophilin mRNA. The transcriptional start sites were identified by primer extension analysis using a ^{32}P-labelled primer having the sequence, 5′-GGGCTCTGTATTTCCCCTAC-3′ (SEQ ID NO:3), and total RNA from day 14 lactating mammary gland. This primer extension assay was adapted from Roussel *et al.* (*DNA Cell Biol.* 14: 777-788 (1995)), which is herein incorporated by reference. Three major labelled products were obtained from this primer extension experiment, suggesting that transcription of *Btn* is initiated from at least three sites, at nucleotides -83, -19 and +1 (FIG. 2) (residues 4611, 4675, 4694 of SEQ ID NO:1) with the most frequently used site at nucleotide T, designated position -83 in FIG. 2.

All three transcription start sites are close to or within the context of, the initiator element 5′-YYA$_{+1}$NWYY-3′ (Javahery *et al., Mol. Cell Biol.* 14:116-127 (1994), herein

incorporated by reference) which can mediate the initiation of transcription in genes lacking conventional TATAA and CCAAT boxes. Two of these sites, at positions -83 and -19, contain one and three mis-matches, respectively, from the consensus sequence, and the site at position -83 is two nucleotides downstream of the more usual A_{+1} start site. The most proximal start site at nucleotide +1 is within a perfect consensus, although paradoxically it does not appear to be the most frequently used.

Although *Btn* does not have conventional TATAA elements, two AT-rich regions, 5′-TGTAAAT-3′ at position -49 (nucleotides 4645-4651 of SEQ ID NO:1), and 5′-TCTAAA-3′ at position -106 (nucleotides 4583-4588 of SEQ ID NO:1) are within 20-25 nucleotides of the two weaker initiator elements. In common with other genes these regions may cooperatively strengthen the initiation of transcription via the TATA- and initiator-binding proteins (Javahery *et al.*, 1994), and this may explain why the start site at position -83 appears to be the most frequently used site. Interestingly this latter site is closest to the sequence 5′-TCTAAA-3′ (position -106 of FIG. 2), which is a characteristic TATA element in many human MHC class I genes (Le Boutellier, *Crit. Rev. Immunol.* 14: 89-129 (1994), herein incorporated by reference). In addition, it should be noted that many of the milk-protein gene promoters have rather similar atypical TATAA boxes, including the sequence 5′-TTTAAAT-3′ in the rat and mouse whey-acidic protein genes (Campbell and Rosen, *Nucl. Acids Res.* 12: 8685-8697 (1984)) and many of the casein genes (Yu-Lee *et al.*, *Nucl. Acids Res.* 14: 1883-1902 (1986)).

Btn also lacks typical CCAAT elements in the expected context approximately 50 nucleotides upstream from TATA sequences (Breathnach and Chambon, *Ann. Rev. Biochem.* 50: 349-383 (1981), herein incorporated by reference). However, there are several potential CCAAT-like elements (double underlined in FIG. 2), including the sequence 5′-ACAAAGT-3′

(nucleotides 4597-4603 of SEQ ID NO:1), which is within 50 nucleotides of the proximal TATA box, and the sequences 5′-CCATTT-3′ and 5′-CATTT-3′ (nucleotides 4546-4551 and 4533-4537 of SEQ ID NO:1, respectively) which are 30-40 nucleotides upstream of the distal TATA box. Of the milk-protein gene promoters sequenced to date, none have conventional CCAAT boxes.

Mapping the 3′-end of mouse butyrophilin mRNA by RT-PCR. The polyadenylation signal sequence in *Btn* was identified by using the RT-PCR to amplify four regions of cDNA around the first potential polyadenylation (poly A) signal sequence (nucleotides 13091 - 13096 of SEQ. ID NO. 1) after the stop codon in *Btn*. Amplified products of the expected size were obtained with primers 5′ of the putative poly(A) signal sequence, while no RT-PCR products were obtained with the primer pairs surrounding this poly (A) signal sequence or encompassing a region 3′ of nucleotide 13,199 (data not shown). These data suggest that the first potential polyadenylation signal in *Btn* is the preferred termination signal and that the 3′ end of the transcripts lies between nucleotides 13,097 and 13,199.

The predicted 5′- and 3′- boundaries of *Btn* lead to estimates of approximately 8.40-8.57 kb for the sizes of the initial gene transcripts, and values of 3.50-3.68 kb for the sizes of the processed mRNAs. These latter estimates are in good agreement with a value of 3.7 kb for the size of mouse butyrophilin mRNA determined by Northern analysis of total RNA from lactating mouse mammary gland using mcDNA3 as an oligonucleotide probe (data not shown).

Sequence analysis of the butyrophilin gene sequence identified single inverted repeats in the 5′ untranslated region (5′ -UTR), and 3′-untranslated region (3′ UTR). Interestingly, the repeat sequence in the 5′-UTR (nucleotides 4807-4814 of SEQ ID NO:1) is the exact complement of the 3′-UTR sequence (nucleotides 12,556-12,563 of SEQ ID NO:1), suggesting

that these sequences play functional roles in the synthesis, stability or regulation of butyrophilin transcripts.

Translation of Mouse Butyrophilin mRNA. The predicted murine butyrophilin amino acid sequence was derived after verification of exon/intron boundaries from the DNA sequences of mouse cDNAs prepared by RT-PCR and the mouse gene sequence. Total RNA was prepared from mouse mammary tissue (day 1 of lactation) (Chomczynski and Sacchi, *Anal. Biochem.* 162: 156-159(1987)) and reverse transcribed into cDNA by incubation with MuMLV reverse transcriptase and random hexamers at 42° C for 15 min, following the protocol described in the Perkin Elmer RT-PCR kit (Perkin Elmer Corp., Branchburg, NJ). The cDNAs, mcDNA 1, 2, 3 and 4 (FIG. 1C) were then prepared by amplifying the indicated regions of DNA by the PCR. The amino acid sequence was predicted from the verified cDNA sequence using the TRANSLATE program from the Wisconsin Genetics Computer Group (GCG) (SEQ ID NO:4). Based on this amino acid sequence, the translational initiation codon, AUG, is predicted to be at nucleotides 4923-4925 of SEQ ID NO: 1. This site is consistent with the predicted location of translation initiation on bovine butyrophilin mRNA and is also within the preferred context for most eukaryotic genes (Kozak, *Nucl. Acids Res.* 15: 8125-8248 (1987)). There are four other potential AUG initiation codons at positions 4650,4743,4765,4776 of SEQ ID NO:1 between the most distal transcriptional initiation site at position 4611 and the predicted translational start site at position 4923 (SEQ ID NO: 1). However, the most distal of these AUG codons is not within the preferred sequence context

and the other three are almost immediately followed, in-frame, by the stop codons TAA, TGA and TAG, respectively. In almost all such latter cases the RNA polymerase continues to scan the mRNA for the next potential AUG start site (Kozak, *Nucl. Acids Res.* 12:3873-3893 (1984)).

Comparison of the DNA sequence of *Btn* with that of butyrophilin cDNA also revealed that, like many other genes in the IgSF (Williams and Barclay, 1988), there is a close correlation between exon organization and functional units of the protein. Thus, exon 1 encodes all of the 5'-UTR and the signal sequence; the location of the signal sequence is designated by the vertical dashed line in FIG. **1B**. Exons 2 and 3 encode the I-set and C-set immunoglobulin-like domains, respectively, and exon 4 encodes the membrane anchor.

Tissue Specific Expression of Mouse Butyrophilin. Previous work has suggested that butyrophilin is specifically expressed in mammary tissue and that expression is maximal during lactation (reviewed in Mather and Jack, 1993). However, this conclusion was based on the use of either relatively insensitive protein and RNA blotting techniques, or immunofluorescence microscopy. Thus, the expression of mRNA in mouse tissues was analyzed with a much more sensitive RNase protection assay.

Riboprobes were prepared from a mouse cDNA, mcDNA3 (FIG. 1C), subcloned into pCR II (Melton *et al.*, *Nucl. Acids Res.* 12:7035-7056 (1984)). For anti-sense riboprobe, the plasmid was linearized by digestion with *Xba*I and the RNA synthesized using SP6 RNA polymerase. For sense riboprobe, the plasmid was linearized by digestion with *Hind*III and the RNA synthesized using T7 RNA polymerase. In each case the RNA was labelled by the inclusion of [α-^{32}P]-dUTP ($\geq$ 800 Ci/mmol) in the reaction mixtures. Total RNA was prepared (Chomczynski and Sacchi, *Anal. Biochem.* 162: 156-159 (1987) from 13 tissues (pancreas,

intestine, spleen, liver, kidney, heart, lung, uterus, ovary, thymus, brain, salivary gland, and mammary) of three Balb/c mice at day 1 of lactation and mammary tissue was pooled from three Balb/c mice at each of several developmental stages (pregnancy, lactation, and involution). Anti-sense or sense riboprobes (2 x 10^6 cpm/sample) were incubated overnight at 47° C with 10 μg total RNA in 30 μl of a hybridization solution (80% (v/v) formamide, 1 mM EDTA, 10 mM sodium citrate and 300 mM sodium acetate, pH 6.4 (Ambion, Austin, TX)). The RNA in each sample was then digested at 37° C, for 30 min with RNAse One (Promega) (5 U/sample) according to the manufacturer's instructions. RNA was recovered following standard procedures (Sambrook *et al.*, 1989) and the samples separated by electrophoresis in a 6% (w/v) denaturing polyacrylamide gel. Radiolabelled riboprobe protected from RNase digestion was detected by exposure of the dried gel to X-ray film.

The size of the anti-sense riboprobe was such that hybridization to butyrophilin mRNA was expected to protect a 625 bp RNA fragment from digestion with RNase. A radiolabelled fragment of the predicted size was only detected in mammary tissue, out of the total of 13 tissues analyzed (data not shown). Analysis of mammary tissue at different developmental stages showed that butyrophilin mRNA is detectable during pregnancy, lactation and involution but not in glands from virgin animals (data not shown). Expression of butyrophilin mRNA appears to increase markedly in the last half of pregnancy and remains at relatively high levels throughout lactation.

Analysis of the Btn *Promoter*. Because *Btn* is specifically expressed in the mammary gland and is associated with the MHC or MHC-related genes (Vernet *et al.*, 1993; Amadou *et al.*, *Genomics* 26: 9-20 (1995)), a search for similarities between the *Btn* promoter and the regulatory elements of mammary-specific or immune system genes was conducted.

Approximately 1.8 kb of *Btn* 5′ flanking sequence, shown in SEQ ID NO:2, was analyzed on either strand by comparison with sequences in the Transcription Factor Data Base (TFD) (Faisst and Meyer, *Nucl. Acids Res.* 20: 3-26 (1992)) and by comparison with the published sequences of the whey-acidic protein, α-lactalbumin, β-lactoglobulin, and casein genes.

Over thirty different classes of potential regulatory elements were identified throughout the sequence. Elements within *Btn* previously shown to be functional in the promoters of other mammary-specific or immune system genes are indicated in FIG. 2. For the sake of clarity, other elements are omitted, unless they are specifically discussed further below.

The mammary-related factors include three potential STAT binding sites identified using the general STAT consensus 5′-TTNC$(N)_3$AA-5′ (Ihle and Kerr, *Trends Genet.* 11:69-74 (1995) herein incorporated by reference) (asterisks, FIG. 2). Additional STAT binding sites can be identified using a broader consensus, TT$(N)_5$AA, based on the work of Lamb *et al.* (*Nucl. Acids Res.* 23: 3283-3289 (1995), herein incorporated by reference) (no asterisks, FIG. 2). Several C/EBP sites were identified, including one between nucleotides -1505 to -1514, which is the imperfect palindrome 5′-ATTAGGTAAT-3′ (SEQ ID NO:5). There appear to be no sites for the pregnancy-specific mammary nuclear factor (5′-TGAT/ATCA-3′, Lee and Oka, 1992, herein incorporated by reference) or the single-stranded nucleic acid binding proteins (various consensus sequences checked, see Altiok and Groner, 1994, herein incorporated by reference). *Btn* contains potential binding sites for NF1, Ets-related proteins (PU.1 site, Klemsz *et al.*, 1990, herein incorporated by reference), heptamer binding sites for Oct 2A, which will bind Oct 1 (Kemler *et al.*, *EMBO J.* 8: 2001-2008 (1989), herein incorporated by reference) and glucocorticoid response elements (½ sites). There are several YY1 sites and at least 11 GMCSF elements (Nimer *et al.*, 1990, herein incorporated by

reference) which will also bind YY1 (Ye *et al.*, *Nucl. Acids Res.* 22: 5672-5678 (1994), herein incorporated by reference). Two negative regulatory elements characterized in the whey-acidic protein gene promoter were identified (Kolb *et al.*, 1994, herein incorporated by reference). These elements (allowing one mis-match each) are within the appropriate context in *Btn*, approximately 270 nucleotides apart in the proximal region of the promoter. Most significantly, no "milk-box" region was found using the consensus sequence of Laird *et al.* (1988), herein incorporated by reference, and no obvious CoREs with composite C/EBP, glucocorticoid response elements and STAT5 sites (Raught *et al.*, 1995, herein incorporated by reference) were identified. Furthermore, comparison of the 5' flanking region of *Btn* with promoters of the casein, whey acidic protein, α-lactalbumin and β-lactoglobulin genes by FASTA (Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85: 2444-2448 (1988) or BLAST (Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990)) showed only limited similarities. The *Btn* promoter therefore appears to have novel features with respect to the regulatory elements of other mammary-specific genes.

The *Btn* promoter lacks the characteristic response elements associated with classical MHC class I and class II genes (Le Boutellier, 1994; Dorn *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 84: 6249-6253 (1987)). However, there are many potential regulatory elements associated with immune system genes including α- and γ-interferon response elements and consensus sequences for TCF-1 (Faisst and Meyer, 1992, herein incorporated by reference) (not shown in Fig. 2), and PU.1 (Klemsz *et al.*, 1990). A repeat element of three GMCSF sites in the proximal promoter was identified which in the same context has been shown to regulate the mitogen-inducible expression of GMCSF in T cells (Nimer *et al.*, 1990).

Accordingly, the mouse *Btn* promoter may be used to direct the expression of desirable proteins in the milk of transgenic animals and to screen for compounds that are mitogenic. As used herein, the term mouse *Btn* promoter means all the sequenced nucleotides from 1 to 4693 of SEQ ID NO:1 or a substantial equivalent. A substantial equivalent is defined as a DNA sequence which enables a DNA fragment containing this sequence to hybridize under stringent conditions to a DNA fragment containing nucleotides 1 to 4693 of SEQ ID NO:1.

In addition to the regulatory elements found in the promoter region of a gene, there is evidence that regulatory sequences involved in tissue-specific expression may also be located in the transcriptional unit of the gene or in 3′ flanking sequences (See, *e.g.*, Charnay *et al.*, Cell 38:251-263 (1984); Gilles *et al.* Cell 33:717-728 (1983)). Thus, the cloned butyrophilin gene may be used as a source of such regulatory sequences. For example, a rDNA construct for expressing a heterologous protein may include a DNA sequence coding for the protein inserted into the first exon of the *Btn* gene. Preferably, the insert is precisely fused to the *Btn* signal sequence for targeting the heterologous protein into the secretory pathway normally involved in secreting butyrophilin into milk.

Example 4: Cloning and Analysis of the Bovine Butyrophilin Promoter

The 5′ untranscribed region of the bovine butyrophilin gene may be cloned from bovine genomic λ phage libraries by standard hybridization methods using the bovine butyrophilin cDNA disclosed in Jack and Mather (1990). By sequencing a clone containing the bovine promoter, herein referred to as BTN1, and comparing the sequence with the mouse promoter sequence, the boundaries of the bovine promoter and regulatory elements contained therein may be identified.

Example 5: Preparation of Synthetic Butyrophilin Promoter Regions

It will be understood by those skilled in the art that an entire butyrophilin promoter may not be necessary to provide a desired biological activity. For example, if production of a heterologous protein in milk is desired, there will be some minimal region or combination of regions within a butyrophilin promoter that is necessary and sufficient to respond to the transcription factors that control expression of the butyrophilin gene in lactating mammary tissue. On the other hand, if the object is to screen for compounds that are mitogenic, there will be some minimal region or combination of regions in a butyrophilin promoter that are necessary and sufficient to direct expression of butyrophilin or other gene in the presence of mitogens. Such minimal promoter regions that are necessary and sufficient to provide a desired biological activity may be identified by deletion analysis using methods well known in the art.

In brief, deletion constructs are prepared containing increasingly smaller portions of λBtn1 or BTN1 operably linked to a reporter gene (*e.g.*, see Example 6 below) and the amounts of reporter gene expression in response to various transcription factors are compared among the deletion constructs. The minimal promoter region(s) of λBtn1 or BTN1 which provide the desired response are then either subcloned from the deletion constructs or constructed from oligonucleotides synthesized on an automated DNA synthesizer. It will be understood that these minimal regions may comprise DNA sequences derived from a butyrophilin gene or their substantial equivalents, as defined above. These minimal promoter regions may then be operably linked to a desired coding sequence and placed in a recombinant expression vector.

Example 6 - Construction of Butyrophilin:hGH Expression Vector

The Allégro® HGH Transient Gene Expression Immunoassay System (Nichols Institute Diagnostics, San Juan Capistrano, CA) may be used to evaluate butyrophilin promoters or promoter regions. In brief, a DNA fragment containing a butyrophilin promoter or minimal promoter region is cloned into the pØGH vector which contains the human growth hormone (hGH) structural gene but lacks a eukaryotic promoter. The resulting fusion plasmid is transfected into a primary mammary cell line, and the hGH secreted into the medium is detected immunologically using a monoclonal antibody-based assay (Nichols Institute Diagnostics, #40-2205). Since the level of secreted hGH is proportional to mRNA levels, promoter activity can be monitored.

In addition to hGH, other reporter genes such as those encoding chloramphenicol acetyltransferase (CAT), green fluorescent proteins or luciferase could be used to evaluate butyrophilin promoter regions. Detection of the products of these reporter genes products is well-known in the art.

Example 7 - Construction of Transgenic Animals

The production of transgenic mammals containing a foreign DNA sequence coding for a desired protein or polypeptide in its germ line is accomplished by procedures well-known in the art. For example, see Rosen, U.S. Patent 5,304,489 (transgenic mice) and Clark *et al.*, U.S. Patent 5,322,775 (transgenic sheep), each of which is herein incorporated by reference. Generally, the process comprises collection of embryos, injection of the DNA into the embryos, transfer of the surviving embryos to surrogate mothers, and screening the offspring for integration and expression of the foreign gene. To construct the transgenic animals embraced by the invention, the injected DNA would be a rDNA construct comprising a

butyrophilin promoter or minimal butyrophilin promoter region(s) operatively linked to a DNA sequence encoding a desired polypeptide. The DNA construct preferably also comprises a signal sequence operatively linked to the DNA sequence encoding the desired polypeptide.

In addition to constructing germ-line transgenic mammals, the invention contemplates the expression of desired coding sequences under the control of a butyrophilin promoter or promoter region(s) in somatic transgenic mammals. As described by Lothar Hennighausen, *J. Cell. Biochem.*, 49: 325-332 (1992), herein incorporated by reference, such animals may be generated by the physical introduction of DNA with a jet injection gun into the mammary epithelial cells of a living lactating animal. See also Furth, P.A. *et al.*, Gene transfer by jet injection into differentiated tissues of living animals and in organ culture, *Mol. Biotechnol.*, 4(2): 121-127 (Oct. 1995), herein incorporated by reference.

Example 8 - Detection of Disease States Associated With Expression of Butyrophilin in Nonlactating Mammals

As discussed above, butyrophilin is a member of the IgSF and its cytoplasmic domain is similar to the cytoplasmic domains in zine-finger proteins. Thus, the expression of butyrophilin in RNA in nonmammary tissue or in mammary tissue of nonlactating animals may be useful for detecting cancer and other disease states in which the butyrophilin promoter is activated.

The principles, preferred embodiments and modes of operation of the present invention have been described in the foregoing specification. The invention which is intended to be protected herein, however, is not to be construed as limited to the particular forms disclosed, since these are to be regarded as illustrative rather than restrictive. Variations and changes may be made by those skilled in the art without departing from the spirit of the invention.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: MATHER Ph.D., IAN H.
OGG Ph.D., SHERRY L.
JACK Ph.D., LUCINDA J.W.
KOMARAGIRI Ph.D., MADHAV V.S.

(ii) TITLE OF INVENTION: THE BUTYROPHILIN GENE PROMOTER AND USES THEREOF

(iii) NUMBER OF SEQUENCES: 5

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: WATSON COLE STEVENS DAVIS, P.L.L.C.
(B) STREET: 1400 K. STREET, N.W.
(C) CITY: WASHINGTON
(D) STATE: D.C.
(E) COUNTRY: USA
(F) ZIP: 20005-2477

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: POULOS III, JAMES A.
(B) REGISTRATION NUMBER: 31714
(C) REFERENCE/DOCKET NUMBER: 6067/JAP69170A

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 202-628-0088
(B) TELEFAX: 202-628-8034

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 14180 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: 129 ES cell genomic library
(B) CLONE: Lambda Btn1

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: 13

(ix) FEATURE:
(A) NAME/KEY: TATA_signal
(B) LOCATION: 4645..4651

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 4611
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "transcription start site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 4675
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "transcription start site"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 4694
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "transcription start site"

(ix) FEATURE:
(A) NAME/KEY: polyA_signal
(B) LOCATION: 13091..13096
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 13097..13199
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "3' end of transcript"

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 4923..4925
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Translational initiation codon"

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 4650..4651
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Translational initiation codon"
/pseudo

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 4743..4745
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Translational initiation codon"
/pseudo

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 4765..4767
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Translational initiation codon"
/pseudo

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 4776..4778
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Translational initiation codon"
/pseudo

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 5002..5520
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron A"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 5872..8332
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron B"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 8615..9485
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron C"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 9636..10206

(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron D"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 10228..10320
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron E"

(ix) FEATURE:
(A) NAME/KEY: intron
(B) LOCATION: 10348..10738
(C) IDENTIFICATION METHOD: experimental
(D) OTHER INFORMATION: /evidence= EXPERIMENTAL
/standard_name= "Intron E"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 4807..4814
(D) OTHER INFORMATION: /standard_name= "Inverted repeat"

(ix) FEATURE:
(A) NAME/KEY: misc_feature
(B) LOCATION: 12556..12563
(D) OTHER INFORMATION: /standard_name= "Inverted repeat"

(ix) FEATURE:
(A) NAME/KEY: promoter
(B) LOCATION: 1..4693

(ix) FEATURE:
(A) NAME/KEY: sig_peptide
(B) LOCATION: 4923..5001

(ix) FEATURE:
(A) NAME/KEY: misc_signal
(B) LOCATION: 11395..11397
(D) OTHER INFORMATION: /standard_name= "Translational stop
site"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

AAGAATAAGA CACACACACA CACACACACA CACACACACA CACACACAGA GAGAGAGACA 60

GAGACAGAGA CAGAGACAGA GAGAGACAGA GAGTAAAATG CATCCAAAAT ATAAATGGTT 120

ATATAGAGAA ATAAAAGCAT AAATGAGTTA TTGTGTTCAG AAACGTGTAG ACAAGTAACA 180

GATAAAGCAG TTTTACATAG ACTACCACAG TGGCTCATGA CATTAATTTC AGTATTCAAG 240

AGACTGAGGC AGGAGGATCA CCATGGCTTC AAGACCATTC TAGACTATAT AGTAAGTTTC 300

AGGCAAGCCT GGGGTACCTT GCCTCAAAAT AAATAAACAA ATAAAATTAA AAAAAAAAAA 360

AGTAATGAGC AGCCATCTTG GCTTCACCTA CCTATAATCC CAGCCTTGGG AGGTGGAGGC 420

AGGAGGATAA TTGCAGTGAT ATCTTCAGCT GTATAGTGTT CTTGGTGCCA CCCTGGGCAA 480

TAAGATACAG TTCTCCACCC TCCTACAAAA GTACAGTTAT ACTTGTTTGC TTTTGAAAGA 540

AGCTATGGAA GTTACCACCC TCAGGTGACT TTTGAGAGAG GGAGGGGAAT TAACCATGCA 600

GACAAGACGG GGCAGAATGT TCTGGGAGAG ACGATGAGCA CCATTATCTG GAGGAGGTGC 660

TTTGAGTAAC CACACCAATT CCGAGTTTGG CCTGCTAGTG GGACAGTGCA GGAAGAGGAA 720

AGAGAAAGGC TTTTCCTTTT CTTCAATCTG TTACCATGGA AACATCTTTG TCATCTACAA 780

AGAACACATT GGAGGAAGGA AAAGAAAAAA AAAAAAACAA ACACAACATG ATCTGTGAAT 840

GAGTCTGTGT TGAGTCTCAT TCAGAGTCAC CCTGGAGAGA TGTGTTACAT GCGGCTGTGG 900

GTCACAGGTT GAACATGCCC AGGACTACCA CAGTGCTTGT CCCTCCCTCC CTTCCAGCTG 960

TCTTTCTCCT GTTTTTATTT TGAGACACAG TGTCCCTGTG TAGCCTAGCT GGTCTGTGTC 1020

TTGTTTTGTA GGCCAGGAAC CCTCCCCCCA CCCCCAACAC ACACACACAC AAACTCAGAG 1080

CAATGCTCCT GCATCAGCCT CCTGTCTGTT ACAAATAGTC ATCTTAATTA CATGTCTTCC 1140

TAGAGCCTAA GGGTTCTGAC GTCAGTGTGG TTCCAAGTCC CAAGTGATGA CAAAAGCCCA 1200

TCTTAAATTA TTCAGTAATC AGTAATATAT TTAAAGATTT ATTATTAGTA GTGTTAGTAT 1260

TAGTAGCAGT ATTTGGTTTT TCGAGATAGG GTTTCTCTGT GTATCCCTGG CTGTCCTGGA 1320

ACTCACTCTG TAGACCAGGC TGGCCTTGAA CTCAGAAATC CGCCTGCCTC TGCCTCCCAA 1380

GTGCTGGGAT TAAAGGGGTG CACCACCACG GCCTTATTTT TCTTATTTTT CATTTGTGTG 1440

TCTGTGTGTG TGCTCAATTG TGGGTTTGTG CACACAAGGG CAAATGCTCC AGCAGTCCAA 1500

AGAGTGTCGG ACCCCTGGGA GCTGGCATCC CAGGTGGCCA TGAACCACTT GACATAGGTT 1560

TCTCTACAAG AGCAGCCATG TTCTTAACTG CTGAAACATC TCTTCAACCC AACTGTTAAT 1620

ATTTTTGTTC TTCACTCAAA TAAGCTAGGA TGGAACATTT AAATGTATTG TATACGTCAT 1680

TTTAAAATAC AAATTGCCCA CAACTGATGA GGCAAGAGTT CGGAGTAAAG TTCTGAAACT 1740

GCTATCTTGA TAATATGCAT TTCCTGTAGG TATGAAGGAG ATGAGTGTGG CATTTCTGGA 1800

TAGCATTCAG ATACACAGGG ATTGTTACAT TCTCAGTCCT CATGCCAGTC CTCAGCATGC 1860

AGAAAATATT AAGCAAAGAA GTAATAAAAT CAGATGTGTG CTTGGGAACA GCACACAACC 1920

TAGCAGCCAA TCAGGCCAAC GAGAGTAGCT AGCTTTGTTC TTAATCATAG TAATTTTTAA 1980

AAATTAAATA AAATTGAGAA GAAATGTTGT CAAAAATATA AAGCACTTAC TTTAAAATTT 2040

GTTTTTAATT TATTTTTCAT TTATGTGGAT GCATGTTAGT CTGTCATGTG TGTGTGGATA 2100

CCTGGGAGGC CCCTGGAATT TTAATTACGG ACCATGTGGG AGCTGGGAGC TGAGCCCGGG 2160

CCCTCTGCAA AAGCAGCCAG CGCTCTTAAC TTCTAAGCCA TCTCCTCAGA CTTCAAATAT 2220

AATACAATTA TTAACTTTAA TTTTTAAAAA GTCCACACAG AAAGAAGACC AGACCTCAAA 2280

ATAGACAGCG ACTGGTCTGG AGGACTCCAG TCTAGATTTT ACCGAGTGGT CAGCTAATCC 2340

AAGAGAAATG CCCAGCCTTG TTACACCACA AAGGTGATAA TGATGATACT AAATTTCACT 2400

AATTTTCATA TAAGCATGTA AGATAATACC TCTTGTTGCT TGCAGGCCTG AAGGACACTC 2460

TTTAGGATGC TATGATCTTT TTAATATTGT AGGGAAGGTC ACTGATAACA TATATATTTA 2520

TGCCCTTTGA GTCAATGGCT TTATTCATGG AACTGGATCA AACAGCATAT CGGGTAGTTA 2580

TCATAGTTAG AAACAAAGAG CTACAAATAA AAATGCATAT CTTTTTTCTA TTTTCTTCTT 2640

CTTCCTCCCT CCTCCTCTTC CTCCTTTTCT TTCTTGTTTC TTCCTTTTTT TTTTTTTTTC 2700

CCAGGCAGGC TCTCACTGCA TTGCTCTAGT TGTCCAGGAA CTTGATCTGT AGATCAGGCT 2760

GGCCTCAAAC TCTTCCTACC TCTGCCTCCC GAATGCTAGG AATAGAAGCA TGTGTTAGCA 2820

TGTCTGTTTT ACTGTGCGTT TAAAAAAAAA AAGTTTATCT TGCCCTTACT GTTTGCTACA 2880

GGCTAGTAAA ACAAACAGAC ATGGTAGATC GATCTATCTG AGTTACAAAA ACAGACCTTC 2940

TTCGAGCCGG ATTCGAACCA GCGACCTAAG GATTTCCAGG TCGAATACTC CTACAGTCCT 3000

CCGCTCTACC AACTGAGCTA TCGAAGGATA CCATGTATAG TGCCTAGCAA AGTCACAAGT 3060

AGCTTAGAGG AGCCACTATG CCTGATTTTA AGCAGTGCTG GGATCTAACT CAGGGCTTCA 3120

TGAATGCTAG ATGGACCTTC TACCAAATGC CAAGTGCATT CTTTTTTTTT TTTTTTTTAA 3180

TTAGGTAATT TCCTCATTTA CATTTCCAAT GCTATCCCAA AAGTCCTCCA TACCCTCCCC 3240

CCGAAGTGCA GTCTTTATAC TAGAAAAAGA ACTAGAAATC TCATAATCTT CGCAAATATA 3300

TGCGTATTAG CTATGCTATG AACTATGCAG GAAAACTTAC TATGAACTTA TCACTATGAA 3360

CTGATATATA TTGTTCTTAA ATTTTATTTT ATATTTATGT ACAGCATAGA AACAATCATT 3420

GATAAAACTG TTTTTTTTCT TTATCTTTGC ATTTTTTCAG TAATAAATGA AAATTCAAAA 3480

CCAAATAAGA AATTGCTGAT CTCATGACTG ATGGCAGGGT GAAGCGCCAG GTCCTTGTGC 3540

AGTTATACCT TGAAGGTGGA CATCCAGTGG ACTCCTGCCA CCCACACCCA CATTCCTGAA 3600

GGTGTCTCAT GGAAAAGATC AGGGAGGGAG AGCTGCAGCC ATTGTGGACT CACTCTTTAG 3660

CTATTCACAG ATGTAATGAC AAAGTAATTT ACTTTCTGGG CTCCTATTCT CTTGCCTGTT 3720

TTGTTTCCAA TACTGTTTGT GTCTAATACT TTTCCAACTT GGCATAATTC AAACAAGGTA 3780

TTAGTAACAT TAGTCTTTTT CTTAAAAGTA ACAAACACCC CACTCTCTTT TGTTTTGTTC 3840

TCCATATGTA GCTCTTGCAA GTCTGGATCT TGCTATGAAG CCCAGACTGG CCTTAAACTT 3900

ACAATGACCC CCGCCTGCGC CCCCCCCTCC CCCCCCATGA ACTTGGGTTA AAAGAACTGA 3960

AGCCACAGAG TTAAATTCAC AGGCTGATGG CCTCATGACT CATTTCAGTT GCTCAAGTCT 4020

TCTTTCTTTT TGTCCCCATT CCCTATATTC GGTACAGCTC TTTAATGCAT ATATCGTTCT 4080

CTTAGGGGAG GAGGATGAAC CCAAACTACC TGACCACTAA TCTGTAGTCC ACATGTTTAA 4140

AAGGCTGCTC CTCCCCCCAC CCCGAATAAA TACACTTGGT CACCTGTGGG CAGGCTTCTC 4200

TAACAGCACA CAGCCTTCTT CCTTCTGAAG AGCTCTCTCT TTGGCCCCGG GGTGACAAGC 4260

AGCCCTTTTC ACTTGATCAC TGTGGCTCTG GCTCCCTTTT CCTCTGGGTC TGTCGAAATC 4320

GGTAGGTGCT TCACTCTCAG CTCAGCTCTC TTTGTCTCTT CTCTGTACTA GGCTTTCTGT 4380

TCCTCAAGCT CTTCAGCTCT GCCTCTCCCC TCTCTCTCAG ACTTTGTCAA GACTGTATGT 4440

ACCTCACGGT GTAACTCCCA GAGATCACCC TCCTGAGAGC TGCTGGGCTT ACAGTTGAGA 4500

AACACACCTT GTCTTTCTCT CCTCCTTCGT TTCATTTCAT GTTCTCCATT TCTACCTCCG 4560

TGGCTTTATC TTCATTATCA CTTCTAAACA CGAATAACAA AGTATCCCAC TCGATTCGAT 4620

TTTACTTTAT TGTTTTATTG TTATTGTAAA TGAGGAGATT TCTTCATTAT CTACAACTGT 4680

GCCTCGCGGC TCCATTCTGG AGGCAGTCGA GGGCTGGAGG ACCAGACGTA CAGAGGAAGG 4740

GTATGGGGCA GGCGCTGTTG TAAAATGGAC TGAAAATGAC CCTGTAGGGG AAATACAGAG 4800

CCCTCCAGGT TGGAAGAAAC TGGTGGAGAA CAGGGCGCTT GCGGAACCCA TAGTTACCTC 4860

CTGACTGTTT CTCTCCCAGC CTGAAGCTCT TGGCGGGCTT CATTGCCCCA GTTAGCTCAG 4920

AGATGGCAGT TCCCACCAAC TCCTGCCTCC TGGTCTGTCT GCTCACCCTC ACTGTCCTAC 4980

AGCTGCCCAC GCTGGATTCG GGTAAGTTTC TGTTCTAGCC TTCTCTTTCT CGCAAAGTTG 5040

GAAGGTCCCT ATAAATAAAT ACCTCTGACC CGGTTTGGCT CCTGGTGGGG GGACCTTCAC 5100

CACAATCCAG TGAGTTCAAA GGAAACCCAC TGCGGGAGGT AATACACACC TGCAATTGCA 5160

GCACCAGGTG GGCCCAGGCA GATTTTTCTG AGGTCAAAAC CTGCCTGGTC TACATAGAGA 5220

TAGCCAGAGC ATCCAGGGTT ACGTAGGGAG CGCCTAGTTG TTTTTCCTTT AAATCAAAGG 5280

AATTGGAACG CTAAGTGTGG TGGTGGTGCA CCCCTGTAAT GATCGCACTT GAGAATTGAG 5340

GGCAAGGAGC TCAAGGCTAC ATAGTGAGCT GGAGGCCACC TTGGGATTTA TGAGATCCAG 5400

TCTGAAAAAT AAACAGAAGA AAAGAAATAG CAGCCACCCC GAGTTCCTTT CTTTACAAGG 5460

AGACTGGCCG GTAGGTCCTC CATCCCAACC CATCGTCCTA TCTGACCTTG TTTATTACAG 5520

CAGCTCCCTT CGATGTGACC GCACCTCAGG AGCCAGTGTT GGCCCTAGTG GGCTCAGATG 5580

CCGAGCTGAC CTGTGGCTTT TCCCCAAACG CGAGCTCAGA ATACATGGAG CTGCTGTGGT 5640

TTCGACAGAC GAGGTCGACA GCGGTACTTC TATACCGGGA TGGCCAGGAG CAGGAGGGCC 5700

AGCAGATGAC GGAGTACCGC GGGAGGGCGA CGCTGGCGAC AGCCGGGCTT CTAGACGGCC 5760

GCGCTACTCT GCTGATCCGA GATGTCAGGG TCTCAGACCA GGGGGAGTAC CGGTGCCTTT 5820

TCAAAGACAA CGACGACTTC GAGGAGGCCG CCGTATACCT CAAAGTGGCT GGTGGGTACA 5880

GACGGGATGT GTCGCCTCGT CACTCCGCGC GGAGACTCTC ACTTTGGGGA GAATCATCGT 5940

GTTCATTCTC CAAATCCAAA CGTATTTTCA CGTTTACGTA AGGTTGTGGT GAGCATCTTA 6000

GATGCTCTGA ACAGCTTCGT GGTTTAATGC CTAAGGATTG ACACCCTAAC AGAGTGTGGT 6060

CCGTTGCTAA AGTTCTTTAT CCACCTCCAA AATGGTTTTA CTCATATTAC TCATGTTGTC 6120

TTCTTCTCCC TGTCTGAGAT CATAAGGAAA GAATACATTG AGCTCTAATT TCCCTCCCTG 6180

TTAGTGATCC AAATCAAGCA AATCTCCCAC TCAGTTTTTC CTACTGTGAA ACCAGAAAGC 6240

TAAATCCAGC AAGAATTTGC AACAAGGAAC TAGATAAGTG AAAAATGCTT TGTTAATGAT 6300

AAAACATCAT GTGCTTATAA AGAAATTCCT ACACCTTAGA CTACTGTGTA TAATACACAT 6360

ATTGCCTTTC TCATTTATTT AGGTATTTTC CTTGCTCCGT TAAGAAAGGA GCTGACATAG 6420

TGTCTCAAAC TCTACAGCTT TAAGAACACT TTGAAGTCCT TTATCAAGTA CTAGGATCAT 6480

TCGTAAAACA ATGAGTTTCC CACACCGGGA GTCGAACCCG GGCCGCCTGG GTGAAAACCA 6540

GGAATCCTAA CCGCTAGACC ATGTGGGAAC TGCTATGCAT ACTTATCTTG CCTCCTCCTC 6600

CCATGTAAGG ATTCCGGACG ATGACACACC TGCTCTTTAG ATGTTGGGAA AGGAATCTAT 6660

CAACTTAACT GTATCCCTAG CTCAAAAATA CATTGCCATG TTTTGCCATA TTTAATGTAC 6720

CAAATATAAC GCTCATATCA TTTTTAGGGA AAGGCATCCT AAAATTATAT AATATATAAA 6780

TTATATAATA TATACATATA CATGAAAATA TGTGTATATA CATATATGTA CATAAATATA 6840

TGTTTATATT CACATATACT TGTGGGTTTG TGTATGATAT TTCAACTGGG AAGTAACACC 6900

CTGTAATTCC AGCAACTGGG AGATACAAGC AGGAAGATTA GAAGTTCAAA ATTGACCTTG 6960

GCTACTTAGA ACCTCGATGT TGTTATTATC TTTTATAAGT GATGGCCATC TTCATAAAAT 7020

GAGTTTAAAT TTTTCATACA CACTCTTTTG AATATGAAGA GCTGTTGAGG TGTTGTTTAA 7080

GGATACTTTT CTAGAGCTCA GAATTTTTCT GTATCCTGTA GAACTGTGAA AGGGGAGAGG 7140

GGAGAAGGAA AAAGAGTTAG GAAAAGAGGA AAGAGGGAAA TAGAGGGAAG GAAAGATTAG 7200

ACTAAATAAA AATGAAAAGT AGCTTTATGT TACCTTTGTT GCTGCTAATT TTCTGTTGCT 7260

ATTTGTTTGT TTGATTTTTT TTTTTTTTTT TTTTTTTTTT TTTTTGCATC AGGGTGTTAC 7320

TATGTAGCTT TGGCTGGCCC CAAACTTGCT ATGCAGACCA GGCTGGCCTA GAATTCATAG 7380

AAAGCCACCT GCTTCACCCT CTCCAGCACT CAGATTAAAG GCCTAGACTA TCACTTTCTC 7440

TGTTGTTATA GAGAAATGGT CTTGAACTCG TTATATAGCA GAGTCTGTTC TAGAACTCCT 7500

GATTCTTCTC CCTCCACCTC CTGAATGCCA TGATTACAGT TGTGTGTCCC CTGTGTTGGT 7560

TTTTGTTGGG GCATAATTCA GTGAGGAAAG ATGAGGTTGA AAACATTTAA GAAAATTCTT 7620

GAGTCTGCAT CCTAGGTAAA GAAAAGTTAA ATTATCAACT GCAAACCTCA AGGGGAAAAA 7680

CAAAACAAAA TTCCAAACTC TGTTCTCACG TATATAGTCT TTTGGGGAGT AGCGGTGTGA 7740

CTCAGTTGGA AGTGTCTTTG CTTAGCATGC ACAAAGCCCT GGGTGGGATC TCTAGCACTG 7800

TCTAAAAATG GTTTGTGGTG GCACGTGTCT CTAAGACCAG CATTTGGGAG GTAGAAGCAA 7860

GAGGATCAGA AGTTCAAGGT CATCTTCGGC TGTTTGAGGT CAGCCTGTGC TACATGATAA 7920

TCTGTCTAAG AAGGAGAATA CTTTCCCACC CATCCTAAAA TATTCTAACC ATAGTCATCT 7980

CATCCTCCAA ATCATGTTAT GCACTTCTAA CCACAGAGGT CTTTCTTGAC TCTAGATCTC 8040

TAAGGCACCT TGCAGCCATT GTTCTTGCTG TTCTTGATAG TTGGACAAAC ACCTCCATGT 8100

CTACTCATCA GACTTTTCCA TATGTTGAAG GATGTGATCT CAATAAGGCT ACATCTCTAA 8160

TAAGATATAA AACATTGTTT TTATTGATCC TTCAGTCTTC ATACAGAAGG ATTAGAGGAA 8220

TCCTACCTAG CCCCAGTCTA CTTTTGCCTC CTCTCTGTCT TTCTCAAGCG AAGATTACCA 8280

TGTTTCCAGG AAAGCATCCA CCAAAAGATT AAGGTCAGTT TCTCTCTAAT AGCTGTGGGT 8340

TCAGATCCTC AAATCAGTAT GACGGTTCAA GAGAATGGAG AAATGGAGCT GGAGTGCACC 8400

TCCTCTGGAT GGTACCCAGA GCCTCAGGTG CAGTGGAGAA CAGGCAACAG AGAGATGCTA 8460

CCATCCACGT CAGAGTCCAA GAAGCATAAT GAGGAAGGCC TGTTCACTGT GGCAGTTTCA 8520

ATGATGATCA GAGACAGCTC CATAAAGAAC ATGTCCTGCT GCATCCAGAA TATCCTCCTT 8580

GGCCAGGGGA AGGAAGTAGA GATCTCCTTA CCAGGTCAGT GGAACTAGTG CTGGGTTCTC 8640

ATGATGACAG AGACTCAGGC CAATATGACT TGGGACCCTG CTCAGAAGGG ACATCATGGC 8700

AAAATTGTTT ACATCTTCCC CTACAGCTCT TGCCTGCTGA CTTAAGGAAA TCCTACCAAC 8760

TAAATTAGAA TAAAGATACT TAGGGCTGGG CTGTATCTCT GAGTGCTTGT GTGGCATGCA 8820

GAAGGTCCTA GGTTTTACCC CTTGGTCTGC ACACACCACC TCCATGCCAG TCTCATAAAA 8880

ATTCCAGAGC TTTATTCCAG AGAAACAGGT GATAGAAAAG CTTTGCCTCT GGAGTCCTTC 8940

CTGACAGGAC CCTTCTCCTT CAATAAGCAA GGAGAATAAA TTATTTTTTC TTCTGATTTG 9000

ACTGTACCCT CTCTGAACAT TTCCTCCCTT CCTTGTTCCA CAATGGAGCT CCATATAGGC 9060

CGCCAAAGAC TGCCAAGTTC CTCCAGGAAC TTTCATCATT TCCAATTTAT TACCTGTGAT 9120

TTAGCAGGAA TCATTCCTTG TTTATTGGCC AATGATTTCC ATCCTATCTT GCATGCAATC 9180

ACCTTTCCTC TTCCTTCCCT ACCTCAGCTA CCTCACTGAT AGTTAACAAG GGATTGCTGT 9240

AAATTTTTAT TTCACATGTT CTGACCCCAA CTGGCTGTTC AGTGTTTGCT TTGGCTCAGG 9300

GTCAAATCTT TCTGGAAAGC TTAGCCTGGA GGGGCAATTC TTGCTGTAGG CAGTGTGAGG 9360

CCACTGAGAG CACTCCCATG TCTGTTCTCC TTTGGTATCA GGAGAGAAGC TGAAGTTGTT 9420

CATTTTCCCA ACCAATGTCC TTTTCGGTTT GTTGTTTGTT CATTTTGTTT TGTGTTGTGT 9480

TTTAGCTCCC TTCGTGCCAA GGCTGACTCC CTGGATAGTA GCTGTGGCTA TCATCTTACT 9540

GGCCTTAGGA TTTCTCACCA TTGGGTCCAT ATTTTTCACT TGGAAACTAT ACAAGGAAAG 9600

ATCCAGTCTG CGGAAGAAGG AATTTGGCTC TAAAGGTAAG TCACTGTCCC CAAGGGCTTT 9660

GTGTCTCGGC TTCCAGGGAA GGTTGAATTC AGGGCTGTTT GGATGACTTC CAACAGGAAG 9720

ATGCTGGATT TTAAAATTCC GAGGTTGGAA GGAACGATAA ACCTTCAAAA GTCACAGGTA 9780

CCTACCTACT GTGAAGAAAA GTGCACGTGA CCCAGGCAAA GTCAAAATCA CCTGGAACTG 9840

TCACTGTGTA CCTGATATTC TTTCACAGCC CAGCTGTAGG CTCTCTGGCC AGTCTAACTC 9900

TGTTGCCCAG GAAGAATGTT CTTATTAAGA TCTAGCCCTG AGTCCTAAGC CAGGAGGAAC 9960

TTCCAGGTGA TTTCTTAGAA ATATTCCGGG GAGTCTCTTG TTAATTAATT AATTTATTTA 10020

ATATTTACAT TTTAGTTTAT TTTGTTTTGC TGGCAGCATT TCTGTTCCTG GTTTGCAGGC 10080

AGAGTTCCTG TCACCAGGGC ACCACAGAGT AAACAGTGTC CCCTTGTGTG TCCCTCATTC 10140

TGGTTTTCCT CCTTCCCCTT TCCCATTATA AAAAAAGCCA TTGACATAAT TTTGTTTGTT 10200

TTCCAGAGAG ACTTCTGGAA GAACTCAGTA AGTATTTTTG TTTTGTTTTG TTTTGTTTTT 10260

TTGTCACGAG ATTTTCTCTC TCCTACTTGT TAACTGATGG TCTCTTTCCT TGCGTTTCAG 10320

GATGCAAAAA GACTGTACTG CATGAAGGTC AGTGGTTCTG AGCTCCTCAC TGCCTCTGAA 10380

GCCCTTCCGT GGGAGTCAAA GACCTGGGAG GCTTGCACTC CAGACTACCT CCTTAGTAAC 10440

AGGATAGAAA CAGGGAAGGT GACAGCGAAT GGTCTCAGCG CTTTCTGGGA GGCATCGCGA 10500

GGACCACTAG CTAGCAGAAG AGCTCCTTTG AGGGATACCG CATTTGATAG TTCTTAAGTC 10560

ATGCCGTAGC TGCCAGTAAG AGATTGGGGC TAGAGAGAAG GACTGCTAGT GAGTGGCCTG 10620

ATAGCTCCCC TACCACAGCT CCTGCAACTC TATTCCACGT CTCTGGGAAG GGGAGATAAT 10680

TCGGGTAGTC TTGATACGGG GACAGGCTGA TGCAGTCTCT CTTTGCCTCC AGTTGACGTG 10740

ACTCTGGATC CAGACACAGC CCACCCCCAC CTCTTCCTGT ATGAAGATTC AAAGTCAGTT 10800

CGATTGGAAG ATTCACGTCA GATCCTGCCT GATAGACCAG AGAGATTTGA CTCCTGGCCC 10860

TGTGTGTTGG GCCGTGAGAC CTTTACTTCA GGGAGACATT ACTGGGAGGT GGAGGTGGGA 10920

GATAGAACTG ACTGGGCCAT TGGTGTGTGT AGGGAGAATG TGGTGAAGAA AGGGTTTGAC 10980

CCCATGACTC CTGATAATGG GTTCTGGGCT GTGGAGTTGT ATGGAAATGG GTACTGGGCC 11040

CTCACCCCAC TCAGGACCTC TCTCCGATTA GCAGGGCCCC CTCGCAGAGT TGGGGTTTTT 11100

CTGGACTATG ACGCAGGAGA CATTTCCTTC TACAACATGA GTAACGGATC TCTTATCTAT 11160

ACTTTCCCTA GCATCTCTTT CTCTGGCCCC CTCCGTCCCT TCTTTTGTCT GTGGTCCTGT 11220

GGTAAAAAGC CCCTGACCAT CTGTTCAACT GCCAATGGGC CTGAGAAAGT CACAGTCATT 11280

GCTAATGTCC AGGACGACAT TCCCTTGTCC CCGCTGGGGG AAGGCTGTAC TTCTGGAGAC 11340

AAAGACACTC TCCATTCTAA ACTGATCCCG TTCTCACCTA GCCAAGCGGC ACCATAACAA 11400

ATATTCCAGC TTCACGACTT TGCCTTCCTT TGACTAATCC CTCATGCCCC GAAGCTTCAG 11460

CTGTTGGCTT CTTGCAGCCC TGCTTCTTCC TGGTGGATGG AGATTAATTC ACATTGGGAA 11520

GGTTAGGTAT GTTGCTGCCA GACAAGGCAG GAAGAAAGGC CATCCTAGTT TGTTTCTGTA 11580

CTAACAGTGG GGAGGAAGAG AGCTGAATCC TAAACTATTT CCAGTGCTCA TATTCCTTCA 11640

GGCCAGAGCC TATAGAGAAG GATTTGGTAC AATCACTCGA GGGATCAAGA GGCAATTAGG 11700

TTGGCATGGA ATTATGGCAG AAACATCTGG AATAGGGGTA TGTGGAATGA CAGGTTTTAG 11760

GTAAGGGAGA ACAAAACCAA ACCATAGGAT GCTGAGAAAG AAAGATCTTG GACTAAACTC 11820

CTAAAAAAGC ACTTAGAGAA GATATGACAG GCAAATGAAG TGAATTTGGT CTAATTTGAT 11880

ACACTTGCCC TGTCCCTAGG GTTTTTCAGT TATATCTCAA TTTTTTTGTT GTTAATTACA 11940

TTTTTGACAG CTTCATACAT GTATATAATG CATTCTAATT ACTCTCACTC TCCTCTATTC 12000

TGTCTTATTT CCCTCCCCTC CCCTCATACC TTCCTTCTTG CTTCAAACCT GGCACACTGA 12060

GTTTAATGGG CTATCATGGG AACATGGATT TAGAGCTTTC CTCTGAGCTC AAGAGAGCAG 12120

GTGTGACTGA ATACAGTGAT TTCCCCTCTC CTACAATCAA TCAGCAGTCA ATAGCTCAGC 12180

TGGGAGGGGT AGGGCCTCAT GAGACTTCCC CTATCAAGGC TAAATGTTGA AAGGGCCAGT 12240

TTTTAGCACC TGTGAGATCA TGATTGCAAG AGCCCAGAAG ACAGCATTGC TCGGTCATTC 12300

TCCCTACCCT TTGGCTTTTC TGGTCTTTTG TCCTCTCTTT CAGGATGTGT CTGAACTCTG 12360

TATCTTAAGT TTTCTATGTC ATGTTCTATA AGATAGAGGA GACTGGCCCT GCTTGTTTGA 12420

GAGCAATGTG AGCAAGCTAG CAAGAGACAG AAAGGAGCGG AGATGAATAG GGGTAGAGAA 12480

AATTTTTAAA CAAACCCTCC AGGTGTGTGT GTGTGTGTGT GTGTCTTCCT CTTTTTTGAC 12540

CTCCCTAAAG GTCAATCCAA CCTCACATTA TTGACTCCAC TAGGTGGGGG TTCTGTGTGT 12600

GTGTGTGTGT GTGTGTGTGT GTGTGTGTGT GTGTTTTAAG ATAGAGGTTT ACTATGTAGC 12660

TTAGGCTGGC TTTGAATTCC TGATCCTCCT GCCTCTACCT TCCAAGTGCT GGAAACATAG 12720

CCACATCCAC CACCCCTATC CAGTCCACCT GGTTTGATTC AGCAACGCTC AGGTAGCATC 12780

GCTGTTTGAT CTGGAGCTGC CAGCTCCCTC GGCCCCCACT GCAATGCTTA ACCCCCTCAC 12840

AGGCACCTTC CCTTGCCTAA CACTGCCATC CTTTTCCACA CTGAGCCATT TGCTCAATGT 12900

AGCCTACCCA GGTATCCTGC TTTCTGGTCC CCAAAGTTAC ACCATGATGC TCAGCACAGC 12960

TGGACAGTTT GTCCCAATTT GTGTGTGTCC TCCTGTTTGT ATGGGACTTC TTTTTGTCAA 13020

TGGCCTGTGT GTGTATCCAA GCTCTTCCAC TTCTATTGTA TTTTTCCGGC TTCTAAAACA 13080

GATGTTACCA AATAAAGAAA GAGAAAGAAA CGAATGTCTG TTTGCTGAAG GCAGCCTCTG 13140

AACTTTTCTT TCTTTATCCC AATAAGAGGG ACTGGATTAA ACCGAAACAG GAATGAGCGC 13200

TGCCTGTCTG GGAAAGTCCT ATTGCAGCAG GGCTGTTCTG TATGGTCCGA GGCTTAGGAC 13260

TGGGAGATTT ACCAGACCAG GCAGAGATGG GAGCTACTCA TGAGGATCAA ACTACCTTCA 13320

AAGAGGCCAC TGTGCTGATG GCTTCCTGCT CTCAGCCTTG TTTCAAAGGC AACTTCATTT 13380

CTATCCCCAC TAAGGTAACT TTGTTGGTGA GTAAACTCCA ACACGGTGCC AGATGTACCA 13440

AGAGGGTGCA GCTCCACAGT AGAGTTCTTG CCCGCCATGC ACCTGCGTGT GTTCCATCCC 13500

TAGCACTGCT TCTGCCCCAC ACATGATTCC TAACAAGTCT CCAAAGACAT GAAAATTGGG 13560

GGATACATTC AAACCACTGC AGGTTCTTTC CTCTATCACC TCATGGGTCC CCGGTGCCCA 13620

GTGTCTTCCT TCTCTTTTTT ATCTCAAACA CTAGCCACCC TATGCAGCTT GTCTTTTACT 13680

GTACTCCTAG GAGAACGGTA TAATTTACCT TTGATTTAAG AGAATTAACT TAATTGAGTG 13740

TGGTGACATG GGCCTGTATT CCCCAGAATT CAGAAGACAG AGGCAAGAGA ATTGTCACAT 13800

ATTTGAAGCC AGCTTGGACT ATATGTCAAT TCAAGGTCAG CCTAAGCTAT ACAGTAACAC 13860

CCTATCTCAT TAAATAAATA AATAAATAAA TGTGTTCATT TTATTCAAAT ATTTTACTTG 13920

TAGAAATCCA CAGAAAATAT AGTCGAAACA TCCTTTCAAA AATTGGTGAG ATGGCTCACC 13980

AGATAAAGAC ACTTACTTGC CAAACCTGAT GACCCGAGTT CAACCCCAGC GACCCACATG 14040

GTGGAGTGAA TTGTCCTCTG ATATCCACAT GTTTGTCATA GATCATGCTC ACCCATACAC 14100

ATATACACAT ACACATGCTA AATATGTTCC ATGTCTAAGA AAGGTAGACT GTTGCATCAC 14160

TGTGTTTAAT GTGTGACAAG 14180

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1750 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: 129 ES Cell Genomic Library
(B) CLONE: Lambda BTN1

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: 13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CGAGCCGGAT TCGAACCAGC GACCTAAGGA TTTCCAGGTC GAATACTCCT ACAGTCCTCC 60

GCTCTACCAA CTGAGCTATC GAAGGATACC ATGTATAGTG CCTAGCAAAG TCACAAGTAG 120

CTTAGAGGAG CCACTATGCC TGATTTTAAG CAGTGCTGGG ATCTAACTCA GGGCTTCATG 180

AATGCTAGAT GGACCTTCTA CCAAATGCCA AGTGCATTCT TTTTTTTTTT TTTTTTAATT 240

AGGTAATTTC CTCATTTACA TTTCCAATGC TATCCCAAAA GTCCTCCATA CCCTCCCCCC 300

GAAGTGCAGT CTTTATACTA GAAAAAGAAC TAGAAATCTC ATAATCTTCG CAAATATATG 360

CGTATTAGCT ATGCTATGAA CTATGCAGGA AAACTTACTA TGAACTTATC ACTATGAACT 420

GATATATATT GTTCTTAAAT TTTATTTTAT ATTTATGTAC AGCATAGAAA CAATCATTGA 480

TAAAACTGTT TTTTTTCTTT ATCTTTGCAT TTTTTCAGTA ATAAATGAAA ATTCAAAACC 540

AAATAAGAAA TTGCTGATCT CATGACTGAT GGCAGGGTGA AGCGCCAGGT CCTTGTGCAG 600

TTATACCTTG AAGGTGGACA TCCAGTGGAC TCCTGCCACC CACACCCACA TTCCTGAAGG 660

TGTCTCATGG AAAAGATCAG GGAGGGAGAG CTGCAGCCAT TGTGGACTCA CTCTTTAGCT 720

ATTCACAGAT GTAATGACAA AGTAATTTAC TTTCTGGGCT CCTATTCTCT TGCCTGTTTT 780

GTTTCCAATA CTGTTTGTGT CTAATACTTT TCCAACTTGG CATAATTCAA ACAAGGTATT 840

AGTAACATTA GTCTTTTTCT TAAAAGTAAC AAACACCCCA CTCTCTTTTG TTTTGTTCTC 900

CATATGTAGC TCTTGCAAGT CTGGATCTTG CTATGAAGCC CAGACTGGCC TTAAACTTAC 960

AATGACCCCC GCCTGCGCCC CCCCCTCCCC CCCCATGAAC TTGGGTTAAA AGAACTGAAG 1020

CCACAGAGTT AAATTCACAG GCTGATGGCC TCATGACTCA TTTCAGTTGC TCAAGTCTTC 1080

TTTCTTTTTG TCCCCATTCC CTATATTCGG TACAGCTCTT TAATGCATAT ATCGTTCTCT 1140

TAGGGGAGGA GGATGAACCC AAACTACCTG ACCACTAATC TGTAGTCCAC ATGTTTAAAA 1200

GGCTGCTCCT CCCCCCACCC CGAATAAATA CACTTGGTCA CCTGTGGGCA GGCTTCTCTA 1260

ACAGCACACA GCCTTCTTCC TTCTGAAGAG CTCTCTCTTT GGCCCCGGGG TGACAAGCAG 1320

CCCTTTTCAC TTGATCACTG TGGCTCTGGC TCCCTTTTCC TCTGGGTCTG TCGAAATCGG 1380

TAGGTGCTTC ACTCTCAGCT CAGCTCTCTT TGTCTCTTCT CTGTACTAGG CTTTCTGTTC 1440

CTCAAGCTCT TCAGCTCTGC CTCTCCCCTC TCTCTCAGAC TTTGTCAAGA CTGTATGTAC 1500

CTCACGGTGT AACTCCCAGA GATCACCCTC CTGAGAGCTG CTGGGCTTAC AGTTGAGAAA 1560

CACACCTTGT CTTTCTCTCC TCCTTCGTTT CATTTCATGT TCTCCATTTC TACCTCCGTG 1620

GCTTTATCTT CATTATCACT TCTAAACACG AATAACAAAG TATCCCACTC GATTCGATTT 1680

TACTTTATTG TTTTATTGTT ATTGTAAATG AGGAGATTTC TTCATTATCT ACAACTGTGC 1740

CTCGCGGCTC 1750

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligodeoxynucleotide"

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: YES

(vi) ORIGINAL SOURCE:
(A) ORGANISM: mus musculus

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: 13

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

GGGCTCTGTA TTTCCCCTAC

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 524 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: not relevant
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(iii) HYPOTHETICAL: NO

(iv) ANTI-SENSE: NO

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Mus musculus

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: 129 ES Cell Genomic Library
(B) CLONE: Lambda BTN1

(viii) POSITION IN GENOME:
(A) CHROMOSOME/SEGMENT: 13

(ix) FEATURE:
(A) NAME/KEY: Domain
(B) LOCATION: 244..270
(D) OTHER INFORMATION: /note= "Membrane anchor domain"

(ix) FEATURE:

(A) NAME/KEY: Peptide
(B) LOCATION: 1..26
(D) OTHER INFORMATION: /note= "Signal Peptide"

(ix) FEATURE:
(A) NAME/KEY: Domain
(B) LOCATION: 27..143
(D) OTHER INFORMATION: /note= "I-set immunoglobulin-like domain"

(ix) FEATURE:
(A) NAME/KEY: Domain
(B) LOCATION: 144..237
(D) OTHER INFORMATION: /note= "C-set Immunoglobulin-like domain"

(ix) FEATURE:
(A) NAME/KEY: Region
(B) LOCATION: 304..469
(D) OTHER INFORMATION: /note= "B30.2 Region"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Ala Val Pro Thr Asn Ser Cys Leu Leu Val Cys Leu Leu Thr Leu
1               5                  10                  15

Thr Val Leu Gln Leu Pro Thr Leu Asp Ser Ala Ala Pro Phe Asp Val
            20                  25                  30

Thr Ala Pro Gln Glu Pro Val Leu Ala Leu Val Gly Ser Asp Ala Glu
        35                  40                  45

Leu Thr Cys Gly Phe Ser Pro Asn Ala Ser Ser Glu Tyr Met Glu Leu
    50                  55                  60

Leu Trp Phe Arg Gln Thr Arg Ser Thr Ala Val Leu Leu Tyr Arg Asp
65                  70                  75                  80

Gly Gln Glu Gln Glu Gly Gln Gln Met Thr Glu Tyr Arg Gly Arg Ala
                85                  90                  95

Thr Leu Ala Thr Ala Gly Leu Leu Asp Gly Arg Ala Thr Leu Leu Ile
            100                 105                 110

Arg Asp Val Arg Val Ser Asp Gln Gly Glu Tyr Arg Cys Leu Phe Lys
        115                 120                 125

Asp Asn Asp Asp Phe Glu Glu Ala Ala Val Tyr Leu Lys Val Ala Ala
    130                 135                 140

Val Gly Ser Asp Pro Gln Ile Ser Met Thr Val Gln Glu Asn Gly Glu
145                 150                 155                 160

Met Glu Leu Glu Cys Thr Ser Ser Gly Trp Tyr Pro Glu Pro Gln Val
                165                 170                 175
```

```
Gln Trp Arg Thr Gly Asn Arg Glu Met Leu Pro Ser Thr Ser Glu Ser
            180                 185                 190

Lys Lys His Asn Glu Glu Gly Leu Phe Thr Val Ala Val Ser Met Met
        195                 200                 205

Ile Arg Asp Ser Ser Ile Lys Asn Met Ser Cys Cys Ile Gln Asn Ile
    210                 215                 220

Leu Leu Gly Gln Gly Lys Glu Val Glu Ile Ser Leu Pro Ala Pro Phe
225                 230                 235                 240

Val Pro Arg Leu Thr Pro Trp Ile Val Ala Val Ala Ile Ile Leu Leu
                245                 250                 255

Ala Leu Gly Phe Leu Thr Ile Gly Ser Ile Phe Phe Thr Trp Lys Leu
            260                 265                 270

Tyr Lys Glu Arg Ser Ser Leu Arg Lys Lys Glu Phe Gly Ser Lys Glu
        275                 280                 285

Arg Leu Leu Glu Glu Leu Arg Cys Lys Lys Thr Val Leu His Glu Val
    290                 295                 300

Asp Val Thr Leu Asp Pro Asp Thr Ala His Pro His Leu Phe Leu Tyr
305                 310                 315                 320

Glu Asp Ser Lys Ser Val Arg Leu Glu Asp Ser Arg Gln Ile Leu Pro
                325                 330                 335

Asp Arg Pro Glu Arg Phe Asp Ser Trp Pro Cys Val Leu Gly Arg Glu
            340                 345                 350

Thr Phe Thr Ser Gly Arg His Tyr Trp Glu Val Glu Val Gly Asp Arg
        355                 360                 365

Thr Asp Trp Ala Ile Gly Val Cys Arg Glu Asn Val Val Lys Lys Gly
    370                 375                 380

Phe Asp Pro Met Thr Pro Asp Asn Gly Phe Trp Ala Val Glu Leu Tyr
385                 390                 395                 400

Gly Asn Gly Tyr Trp Ala Leu Thr Pro Leu Arg Thr Ser Leu Arg Leu
                405                 410                 415

Ala Gly Pro Pro Arg Arg Val Gly Val Phe Leu Asp Tyr Asp Ala Gly
            420                 425                 430

Asp Ile Ser Phe Tyr Asn Met Ser Asn Gly Ser Leu Ile Tyr Thr Phe
        435                 440                 445

Pro Ser Ile Ser Phe Ser Gly Pro Leu Arg Pro Phe Phe Cys Leu Trp
    450                 455                 460

Ser Cys Gly Lys Lys Pro Leu Thr Ile Cys Ser Thr Ala Asn Gly Pro
465                 470                 475                 480
```

Glu Lys Val Thr Val Ile Ala Asn Val Gln Asp Asp Ile Pro Leu Ser
485 490 495

Pro Leu Gly Glu Gly Cys Thr Ser Gly Asp Lys Asp Thr Leu His Ser
500 505 510

Lys Leu Ile Pro Phe Ser Pro Ser Gln Ala Ala Pro
515 520

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other nucleic acid
(A) DESCRIPTION: /desc = "oligodeoxynucleotide"

(iii) HYPOTHETICAL: YES

(iv) ANTI-SENSE: NO

(ix) FEATURE:
(A) NAME/KEY: protein_bind
(B) LOCATION: 1..10
(D) OTHER INFORMATION: /bound_moiety= "Transcription factor"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATTAGGTAAT 10

We claim:

1. A purified and isolated DNA fragment comprising a DNA sequence having the biological activity of a butyrophilin promoter.

2. The DNA fragment of claim 1, wherein the DNA sequence comprises at least one minimal promoter region from the mouse *Btn* promoter, the bovine *BTN* promoter, or their substantial equivalents.

3. The DNA fragment of claim 2, wherein the minimal promoter region is from the mouse *Btn* promoter.

4. The DNA fragment of claim 1, wherein the DNA sequence is selected from the group consisting of:

(a) nucleotides 1 to 4693 of SEQ ID NO:1; and

(b) DNA sequences which are substantial equivalents of the sequences defined in (a).

5. The DNA fragment of claim 4, wherein the DNA sequence further comprises nucleotides 4694-4922 of SEQ ID NO:1, wherein the nucleotides 4694-4922 are contiguous with the nucleotides 1 to 4693.

6. The DNA fragment of claim 5, wherein the DNA sequence further comprises nucleotides 4923-5001 of SEQ ID NO:1, wherein the nucleotides 4923-5001 are contiguous with the nucleotides 4694-4922.

7. The DNA fragment of claim 6, wherein the DNA sequence further comprises nucleotides 5002-14180 of SEQ ID NO:1, or its complementary sequence, wherein the nucleotides 5002-14180 are contiguous with the nucleotides 4923-5001.

8. The DNA fragment of claim 1, wherein the DNA sequence is the bovine *BTN* promoter.

9. A rDNA construct for expressing a polypeptide in the mammary gland of a mammal, the rDNA construct comprising:

(a) a first DNA sequence having the biological activity of the butyrophilin promoter; and

(b) a second DNA sequence encoding the polypeptide operatively linked to the first DNA sequence.

10. The rDNA construct of claim 7, wherein the first DNA sequence comprises at least one minimal butyrophilin promoter region.

11. The rDNA construct of claim 10, wherein the minimal promoter region is from the mouse *Btn* promoter, the bovine *BTN* promoter, or their substantial equivalents.

12. The rDNA construct of claim 9, wherein the first DNA sequence is selected from the group consisting of:

(a) a DNA sequence comprising nucleotides 1 to 4693 of SEQ ID NO:1; and

(b) DNA sequences which are substantial equivalents of the sequence defined in (a).

13. The rDNA construct of claim 12, wherein the first DNA sequence further comprises nucleotides 4694-4922 of SEQ ID NO:1, or its complementary sequence, contiguous with the nucleotides 1 to 4693.

14. The rDNA construct of claim 9, further comprising a third DNA sequence encoding a protein signal sequence operatively linked between first and second DNA sequences.

15. The rDNA construct of claim 14, wherein the signal sequence is a milk protein signal sequence and the third DNA sequence is fused to the second DNA sequence.

16. The rDNA construct of claim 15, wherein the third DNA sequence is selected from the group consisting of:

(a) a DNA sequence comprising nucleotides 4923-5001 of SEQ ID NO:1, or its complementary sequence, and

(b) DNA sequences which are substantial equivalents of the sequences defined in (a).

17. A transgenic mammal containing a rDNA construct in at least its mammary epithelial cells, the rDNA construct comprising

(a) a first DNA sequence having the biological activity of a butyrophilin promoter; and

(b) a second DNA sequence encoding a polypeptide operatively linked to the first DNA sequence, the rDNA construct being integrated in such a way that the second DNA sequence is expressed in the mammary gland of the transgenic mammal and the polypeptide is present in the milk of the mammal.

18. The transgenic mammal of claim 17, wherein the first DNA sequence comprises at least one minimal butyrophilin promoter region.

19. The transgenic mammal of claim 18, wherein the minimal promoter region is from the mouse *Btn* promoter, the bovine *BTN* promoter, or their substantial equivalents.

20. The transgenic mammal of claim 17, wherein the first DNA sequence is selected from the group consisting of:

(a) a DNA sequence comprising nucleotides 1 to 4693 of SEQ ID NO:1; and

(b) DNA sequences which are substantial equivalents of the sequence defined in (a).

21. The transgenic mammal of claim 20, wherein the first DNA sequence further comprises nucleotides 4694-4922 of SEQ ID NO:1, or its complementary strand, contiguous with the nucleotides 1 to 4693.

22. The transgenic mammal of claim 17, wherein the rDNA construct further comprises a third DNA sequence encoding a signal sequence operatively linked between the first and second DNA sequences.

23. The transgenic mammal of claim 22, wherein the signal sequence is a milk protein signal sequence and the third DNA sequence is fused to the second DNA sequence.

24. The transgenic mammal of claim 17, wherein the rDNA construct is also present in the germ cells and all the somatic cells of the transgenic mammal.

25. A method of producing a polypeptide comprising the steps of

(a) producing milk in a transgenic mammal, the mammal containing a rDNA construct in at least its mammary epithelial cells, the rDNA construct comprising

(i) a first DNA sequence having the biological activity of a *Btn* promoter; and

(ii) a second DNA sequence encoding the polypeptide operatively linked to the first DNA sequence; the rDNA construct being integrated in such a way that the second DNA sequence is expressed in the mammary gland of the transgenic mammal and the polypeptide is present in the milk; and

(b) collecting the milk produced in step (a).

26. The method of claim 25, further comprising:

(c) removing the polypeptide from the collected milk.

27. The method of claim 25, wherein the rDNA construct is also present in the germ cells and all the somatic cells of the transgenic mammal.

28. A method for detecting a disease state associated with activation of a *Btn* promoter in nonlactating mammals comprising detecting expression of butyrophilin mRNA or protein in a tissue of a nonlactating mammal.

29. The method of claim 28 wherein the disease state is breast cancer and the tissue is breast cancer.

30. A method for testing the carcinogenicity of a substance comprising comparing the level of expression of a reporter gene in a recombinant cell in the presence of the substance with the level of expression of the reporter gene in the recombinant cell in the absence of the substance, the recombinant cell containing a rDNA construct comprising

(a) the first DNA sequence having the biological activity of a butyrophilin promoter; and

(b) a second DNA sequence encoding the reporter gene operatively linked to the first DNA sequence.

31. A purified and isolated DNA fragment comprising a DNA sequence coding for a polypeptide having the amino sequence of SEQ ID NO:4.

32. A purified and isolated DNA fragment comprising a DNA sequence encoding mouse butyrophilin, wherein said DNA sequence comprises nucleotides 4694-13199.

33. A purified and isolated DNA fragment comprising a DNA sequence coding for the promoter and transcriptional unit of the mouse butyrophilin gene, said DNA sequence obtained by a process comprising the steps of:

(a) growing λBtn1 (ATCC Deposit No. 97513) on a host bacteria strain to generate a lysate of λBtn phage particles;

(b) concentrating the λBtn1 phage particles;

(c) extracting λBtn1 DNA from the concentrated phage particles; and

(d) sequencing the extracted λBtn1 DNA.

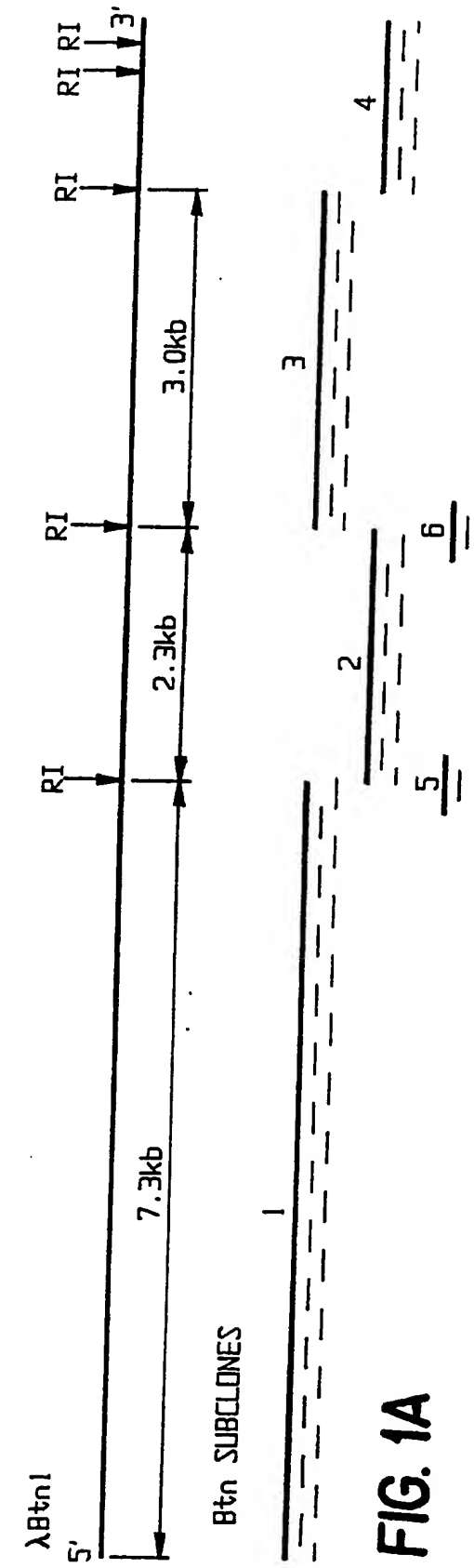

FIG. 1A

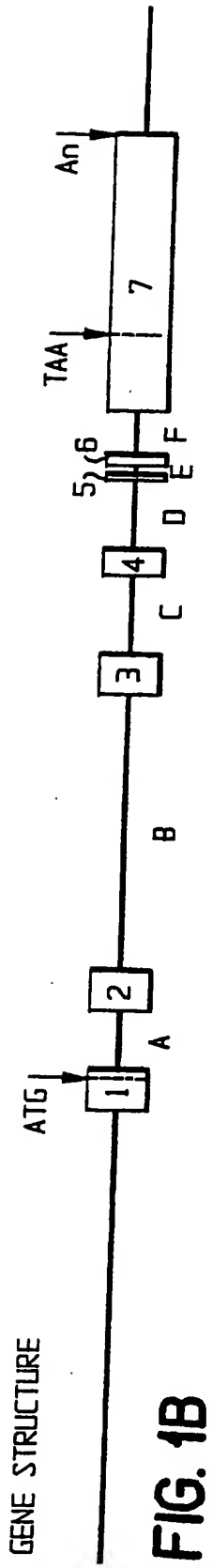

FIG. 1B

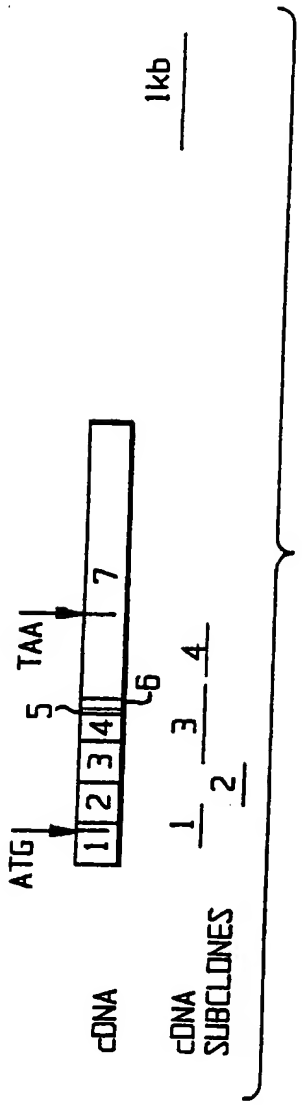

FIG. 1C

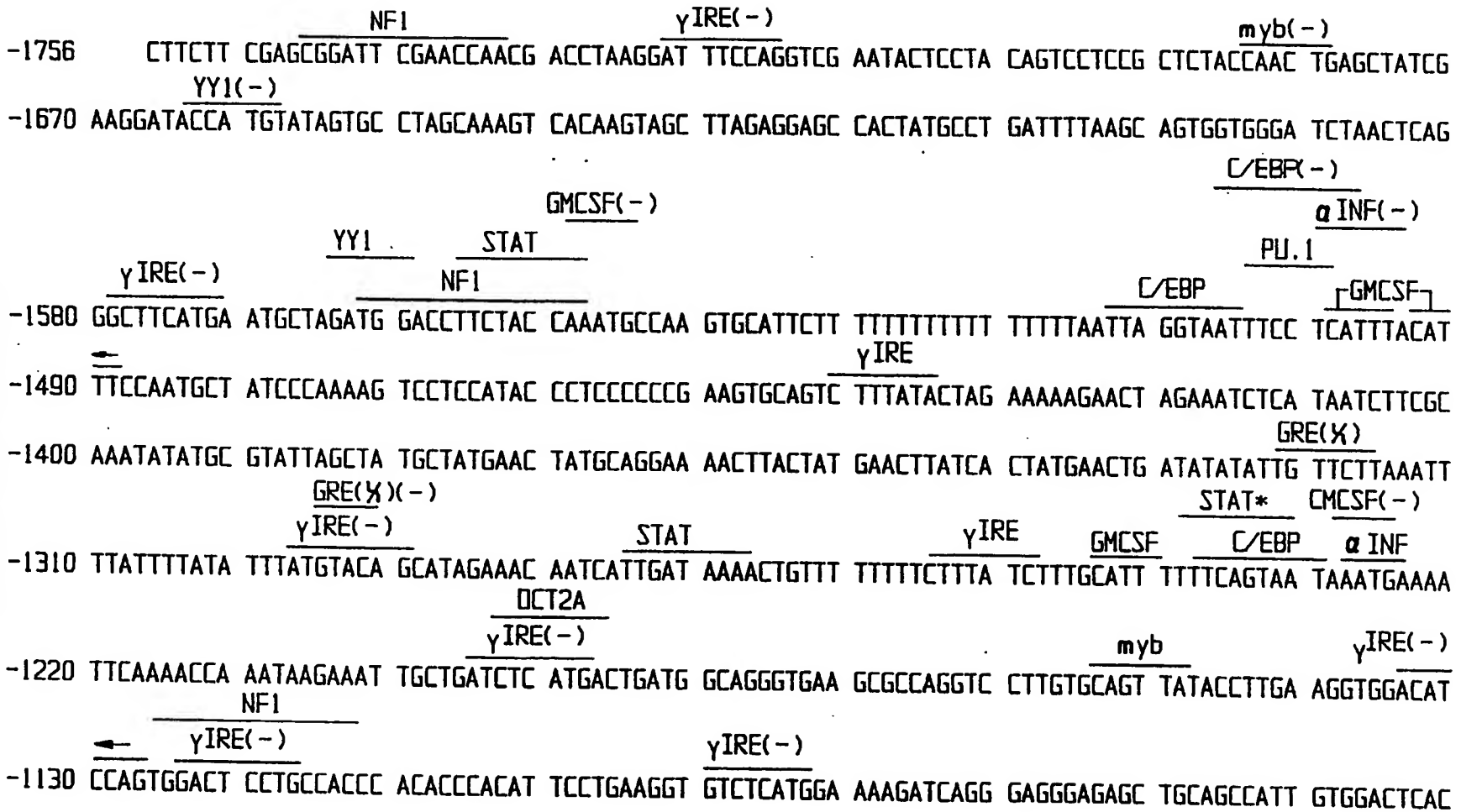
NF1
γIRE(-)
myb(-)
-1756 CTTCTT CGAGCGGATT CGAACCAACG ACCTAAGGAT TTCCAGGTCG AATACTCCTA CAGTCCTCCG CTCTACCAAC TGAGCTATCG
YY1(-)
-1670 AAGGATACCA TGTATAGTGC CTAGCAAAGT CACAAGTAGC TTAGAGGAGC CACTATGCCT GATTTTAAGC AGTGGTGGGA TCTAACTCAG
C/EBP(-)
GMCSF(-)
αINF(-)
YY1
STAT
PU.1
γIRE(-)
NF1
C/EBP
GMCSF
-1580 GGCTTCATGA ATGCTAGATG GACCTTCTAC CAAATGCCAA GTGCATTCTT TTTTTTTTTT TTTTTAATTA GGTAATTTCC TCATTTACAT
γIRE
-1490 TTCCAATGCT ATCCCAAAAG TCCTCCATAC CCTCCCCCCG AAGTGCAGTC TTTATACTAG AAAAAGAACT AGAAATCTCA TAATCTTCGC
GRE(X)
-1400 AAATATATGC GTATTAGCTA TGCTATGAAC TATGCAGGAA AACTTACTAT GAACTTATCA CTATGAACTG ATATATATTG TTCTTAAATT
GRE(X)(-)
STAT*
CMCSF(-)
γIRE(-)
STAT
γIRE
GMCSF
C/EBP
αINF
-1310 TTATTTTATA TTTATGTACA GCATAGAAAC AATCATTGAT AAAACTGTTT TTTTTCTTTA TCTTTGCATT TTTTCAGTAA TAAATGAAAA
OCT2A
γIRE(-)
myb
γIRE(-)
-1220 TTCAAAACCA AATAAGAAAT TGCTGATCTC ATGACTGATG GCAGGGTGAA GCGCCAGGTC CTTGTGCAGT TATACCTTGA AGGTGGACAT
NF1
γIRE(-)
γIRE(-)
-1130 CCAGTGGACT CCTGCCACCC ACACCCACAT TCCTGAAGGT GTCTCATGGA AAAGATCAGG GAGGGAGAGC TGCAGCCATT GTGGACTCAC

FIG. 2A

γIRE(-)
γIRE E2A GMCSF(-)
-1040 TCTTTAGCTA TTCACAGATG TAATGACAAA GTAATTTACT TTCTGGGCTC CTATTCTCTT GCCTGTTTTG TTTCCAATAC TGTTTGTGTC
NF1(-) GMCFS
STAT STAT γIRE STAT*
-950 TAATACTTTT CCAACTTGGC ATAATTCAAA CAAGGTATTA GTAACATTAG TCTTTTTCTT AAAAGTAACA AACACCCCAC TCTCTTTTGT
GRE(X) YY1(-) γIRE AP2
ERE(X)(-) SP1
-860 TTTGTTCTCC ATATGTAGCT CTTGCAAGTC TGGATCTTGC TATGAAGCCC AGACTGGCCT TAAACTTACA ATGACCCCCG CCTGCGCCCC
γIRE γIRE myb
STAT DCT2A αINF(-)
AP2 SITES STAT γIRE(-) GMCSF
-770 CCCCTCCCCC CCCATGAACT TGGGTTAAAA GAACTGAAGC CACAGAGTTA AATTCACAGG CTGATGGCCT CATGACTCAT TTCAGTTGCT
GRE(X) GMCSF(-)
-680 CAAGTCTTCT TTCTTTTTGT CCCCATTCCC TATATTCGGT ACAGCTCTTT AATGCATATA TCGTTCTCTT AGGGGAGGAG GATGAACCCA
ERE(X)(-) γIRE E-BOX
AP2 ERE(X)
-590 AACTACCTGA CCACTAATCT GTAGTCCACA TGTTTAAAAG GCTGCTCCTC CCCCCACCCC GAATAAATAC ACTTGGTCAC CTGTGGGCAG
WAP NRE ? AP2 αINF(-)
-500 GCTTCTCTAA CAGCACACAG CCTTCTTCCT TCTGAAGAGC TCTCTCTTTG GCCCCGGGGT GACAAGCAGC CCTTTTCACT TGATCACTGT
PU.1(-) C/EBP GRE(X)(-)
-410 GGCTCTGGCT CCCTTTTCCT CTGGGTCTGT CGAAATCGGT AGGTGCTTCA CTCTCAGCTC AGCTCTCTTT GTCTCTTCTC TGTACTAGGC

FIG. 2B

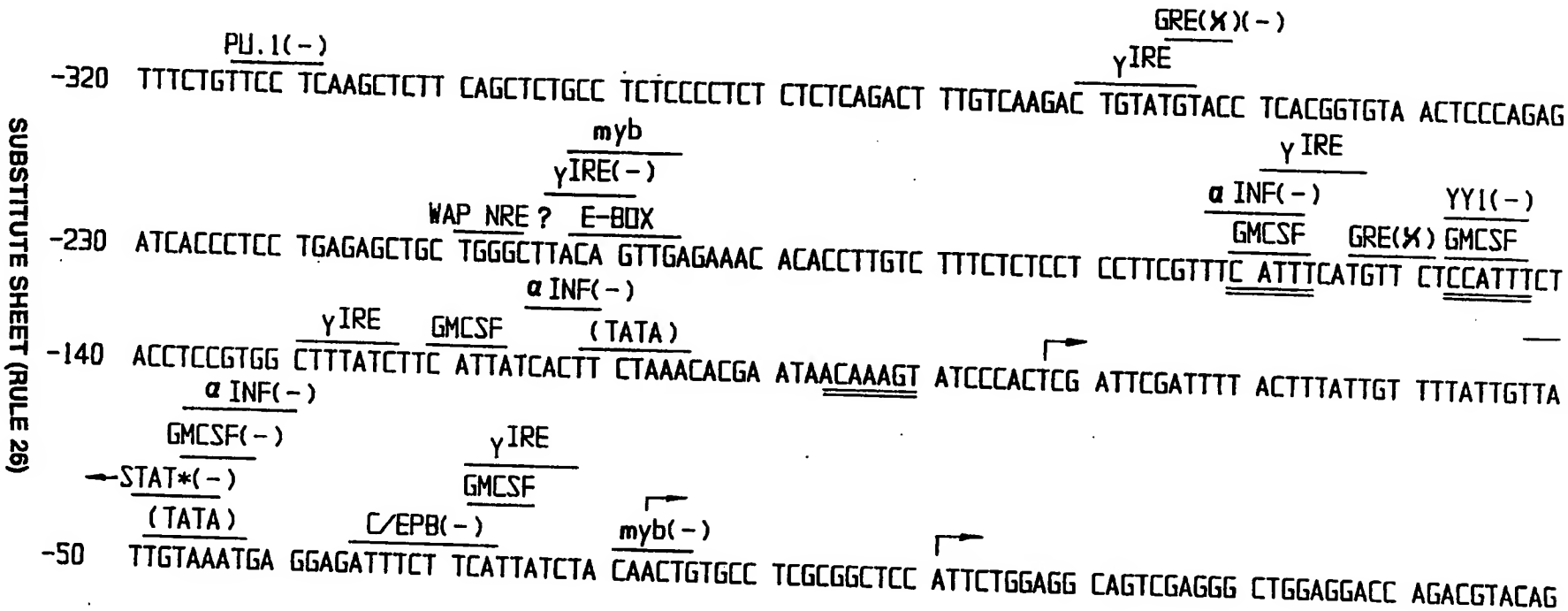
PU.1(-)
GRE(X)(-)
γIRE
-320 TTTCTGTTCC TCAAGCTCTT CAGCTCTGCC TCTCCCCTCT CTCTCAGACT TTGTCAAGAC TGTATGTACC TCACGGTGTA ACTCCCAGAG
myb
γIRE(-)
γIRE
αINF(-)
YY1(-)
WAP NRE? E-BOX
GMCSF
GRE(X) GMCSF
-230 ATCACCCTCC TGAGAGCTGC TGGGCTTACA GTTGAGAAAC ACACCTTGTC TTTCTCTCCT CCTTCGTTTC ATTTCATGTT CTCCATTTCT
αINF(-)
γIRE
GMCSF
(TATA)
-140 ACCTCCGTGG CTTTATCTTC ATTATCACTT CTAAACACGA ATAACAAAGT ATCCCACTCG ATTCGATTTT ACTTTATTGT TTTATTGTTA
αINF(-)
GMCSF(-)
γIRE
STAT*(-)
GMCSF
(TATA)
C/EPB(-)
myb(-)
-50 TTGTAAATGA GGAGATTTCT TCATTATCTA CAACTGTGCC TCGCGGCTCC ATTCTGGAGG CAGTCGAGGG CTGGAGGACC AGACGTACAG

FIG. 2C

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/12933

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :A61K 49/00; C12P 21/06; C12N 5/00, 15/00; C07H 19/00, 21/04
US CL :424/9.1; 435/69.1; 320.1; 536/23.1, 24.1; 800/2
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 424/9.1; 435/69.1; 320.1; 536/23.1, 24.1; 800/2

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, Chemical Abstracts

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	JACK, L. J.W. et al. Cloning and Analysis of cDNA Encoding Bovine Butyrophilin, an Apical Glycoprotein Expressed in Mammary Tissue and Secreted in Association with the Milk-Fat Globule Membrane During Lactation. Journal of Biological Chemistry. 25 August 1993. Vol. 265. No. 24. pages 14481-14486, especially page 14483 and 14484.	1-33
Y	TAYLOR, M.R. et al. Cloning and Sequence Analysis of Human Butyrophilin Reveals a Potential Receptor Function. Biochemic et Biophysica Acta. 1996. Volume 1306, pages 1-4, especially page 2.	1-33

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search 11 SEPTEMBER 1997	Date of mailing of the international search report 09 OCT 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DEBORAH CROUCH, PH.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*